

HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

UTILITY PATENT APPLICATION TRANSMITTAL

*(Only for new nonprovisional applications under
37 C.F.R. 1.53(b))*

Attorney Docket No.

2825.1016-001

First Named Inventor or
Application Identifier

Jian-Bing Fan

Express Mail Label No.

EL552279893US

Title of
Invention

UNIVERSAL ARRAYS

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

ADDRESS TO:

Assistant Commissioner for Patents
Box Patent Application
Washington, D.C. 20231

1. ☐ Fee Transmittal Form
(Submit an original, and a duplicate for fee processing)

6. ☐ Microfiche Computer Program *(Appendix)*

2. ☒ Specification **[Total Pages 60]**
(preferred arrangement set forth below)

7. ☐ Nucleotide and/or Amino Acid Sequence Submission
(if applicable, all necessary)

- Descriptive title of the invention
- Cross References to Related Applications
- Statement Regarding Fed sponsored R & D
- Reference to microfiche Appendix
- Background of the Invention
- Summary of the Invention
- Brief Description of the Drawings
- Detailed Description
- Claim(s)
- Abstract of the Disclosure

a. ☐ Computer Readable Copy

b. ☐ Paper Copy (identical to computer copy)

☐ Pages

c. ☐ Statement verifying identity of above copies

3. ☒ Drawing(s) (35 U.S.C. 113) **[Total Sheets 9]**
☒ Formal ☐ Informal

ACCOMPANYING APPLICATION PARTS

8. ☐ Assignment Papers (cover sheet & documents)

9. ☐ 37 C.F.R. 3.73(b) Statement ☐ Power of Attorney
(when there is an assignee)

10. ☐ English Translation Document *(if applicable)*

11. ☐ Information Disclosure Statement (IDS)/PTO-1449 ☐ Copies of IDS Citations

12. ☐ Preliminary Amendment

13. ☒ Return Receipt Postcard (MPEP 503)
(Should be specifically itemized)

14. ☐ Small Entity Statement(s) ☐ Statement filed in prior application, status still proper and desired

15. ☐ Certified Copy of Priority Document(s)
(if foreign priority is claimed)

16. ☐ Other: _____

4. ☐ Oath or Declaration/POA **[Total Pages []]**

a. ☐ Newly executed (original or copy)

b. ☐ Copy from a prior application (37 C.F.R. 1.63(d))
(for continuation/divisional with Box 17 completed)
[NOTE Box 5 below]

i. ☐ DELETION OF INVENTOR(S)
Signed statement attached deleting inventor(s) named in the prior application, see 37 C.F.R. 1.63(d)(2) and 1.33(b).

5. ☐ Incorporation By Reference *(useable if Box 4b is checked)*
The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.

17. If a **CONTINUING APPLICATION**, check appropriate box and supply the requisite information:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No.:

Prior application information: Examiner:

Group Art Unit:

18. CORRESPONDENCE ADDRESS

NAME	Doreen M. Hogle, Esq.				
	HAMILTON, BROOK, SMITH & REYNOLDS, P.C.				
ADDRESS	Two Militia Drive				
CITY	Lexington	STATE	MA	ZIP CODE	02421-4799
COUNTRY	USA	TELEPHONE	(781) 861-6240	FAX	(781) 861-9540

Signature	Lisa M. Treannie	Date	3/27/00
Submitted by Typed or Printed Name	Lisa M. Treannie	Reg. Number	41,368

DMH/LMT/pdd
3/27/00

-1-

Date: <u>3/27/00</u>	Express Mail Label No. <u>EL552279893US</u>
----------------------	---

Inventors: Jian-Bing Fan, Joel N. Hirschhorn, Xiaohua Huang, Paul Kaplan, Eric S. Lander, David Lockhart, Thomas Ryder and Pamela Sklar

Attorney's Docket No.: 2825.1016-001

UNIVERSAL ARRAYS

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial Nos. 60/126,473, filed March 26, 1999, and 60/140,359, filed June 23, 1999, the entire
5 teachings of both of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

Obtaining genotype information on thousands of polymorphic markers in a highly parallel fashion is becoming an increasingly important task in mapping disease loci, in identifying quantitative trait loci, in diagnosing tumor loss of heterozygosity,
10 and in performing linkage studies. A currently available method for simultaneously obtaining large numbers of polymorphic marker genotypes involves hybridization to allele specific probes on high density oligonucleotide arrays. In order to practice the method, redundant sets of hybridization probes, typically twenty or more, are used to score each marker. A high degree of redundancy is required, however, to reduce the
15 noise and achieve an acceptable level of accuracy. Even this level of redundancy is often insufficient to unambiguously score heterozygotes or to quantitatively determine allele frequency in a population. Thus, there is a need in the art for more reliable and better quantitative methods to identify genotypes at polymorphic markers.

SUMMARY OF THE INVENTION

An array of oligonucleotide tags attached to a solid substrate is disclosed, along with locus-specific tagged oligonucleotides. The array and the locus-specific tagged oligonucleotides are particularly useful in genotyping using single base extension
5 reactions. When used together, the array and the locus-specific tagged oligonucleotides serve as a "universal chip" system for use in genotyping, wherein by using different sets of locus-specific tagged oligonucleotides the system can be tailored to any desired genotyping application. For example, it is an object of the present invention to provide a method to aid in determining a ratio of alleles at a polymorphic locus. It is another
10 object of the invention to provide a set of primers for use in determining a ratio of nucleotides present at a polymorphic locus.

Thus, in one embodiment the invention relates to an array comprising one or more oligonucleotide tags fixed to a solid substrate, wherein each oligonucleotide tag comprises a unique known arbitrary nucleotide sequence of sufficient length to
15 hybridize to a locus-specific tagged oligonucleotide, wherein the locus-specific tagged oligonucleotide has at its first end nucleotide sequence which hybridizes to, e.g., is complementary to, the arbitrary sequence of the oligonucleotide tag, and wherein the locus-specific tagged oligonucleotide has at a second end nucleotide sequence complementary to target polynucleotide sequence in a sample.

20 In one embodiment, the invention relates to a kit comprising an array comprising one or more oligonucleotide tags fixed to a solid substrate, wherein each oligonucleotide tag comprises a unique known arbitrary nucleotide sequence of sufficient length to hybridize to a locus-specific tagged oligonucleotide, and one or more locus-specific tagged oligonucleotides, wherein each locus-specific tagged
25 oligonucleotide has at its first (5') end nucleotide sequence which hybridizes to, e.g., is complementary to, the arbitrary sequence of a corresponding oligonucleotide tag on the array, and has at its second (3') end nucleotide sequence complementary to target polynucleotide sequence in a sample.

The invention further relates to a method of genotyping a nucleic acid sample at one or more loci, comprising the steps of obtaining a nucleic acid sample to be tested; combining the nucleic acid sample with one or more locus-specific tagged oligonucleotides under conditions suitable for hybridization of the nucleic acid sample to one or more locus-specific tagged oligonucleotides, wherein each locus-specific tagged oligonucleotide comprises a nucleotide sequence capable of hybridizing to a complementary sequence in an oligonucleotide tag and a nucleotide sequence complementary to the nucleotide sequence 5' of a nucleotide to be queried in the sample, thereby creating an amplification product-locus-specific tagged oligonucleotide complex; subjecting the complex to a single base extension reaction, wherein the reaction results in the addition of a labeled ddNTP to the locus-specific tagged oligonucleotide, and wherein each type of ddNTP has a label that can be distinguished from the label of the other three types of ddNTPs; contacting the complex with an oligonucleotide array comprising one or more oligonucleotide tags fixed to a solid substrate under suitable hybridization conditions, wherein each oligonucleotide tag comprises a unique arbitrary sequence complementary and of sufficient length to hybridize to a complementary sequence in a locus-specific tagged oligonucleotide, whereby the complex hybridizes to a specific oligonucleotide tag on the array; and assaying the array to determine the labeled ddNTPs present in the complex hybridized to one or more oligonucleotide tags, thereby determining the genotype of the queried nucleotide in the sample. In one embodiment the nucleic acid sample to be tested is amplified.

In one embodiment a method is provided to aid in determining a ratio of alleles at a polymorphic locus in a sample. A pair of primers is used to amplify a region of a nucleic acid in a sample. In one embodiment, the region comprises a polymorphic locus, and an amplified nucleic acid product is formed which comprises the polymorphic locus. The amplified nucleic acid product is used as a template in a single base extension reaction with an extension primer, forming a labeled extension primer. The extension primer (also called a locus-specific tagged oligonucleotide herein)

comprises a 3' portion and a 5' portion. The 3' portion is complementary to the amplified nucleic acid product and terminates one nucleotide 5' to the polymorphic locus. The 5' portion is not complementary to the amplified nucleic acid product. A labeled dideoxynucleotide which is complementary to the polymorphic locus is coupled to the 3' end of the extension primer. Each type of dideoxynucleotide present in the reaction bears a distinct label. The 5' portion of the extension primer is hybridized to one or more probes (also called oligonucleotide tags herein) which are immobilized to known locations on a solid support. The probes comprise a nucleotide sequence which is complementary to the 5' portion of the extension primer.

Also provided by the present invention is a set of primers for use in determining a ratio of nucleotides present at a polymorphic locus. The set includes a pair of amplification primers and an extension primer. The pair of primers prime synthesis of a region of double stranded nucleic acid which comprises a polymorphic locus. The extension primer comprises a 3' portion which is complementary to a portion of the region of double stranded nucleic acid and a 5' portion which is not complementary to the region of double stranded nucleic acid. The extension primer terminates one nucleotide 5' to the polymorphic locus. Examples of primers according to the invention are shown in Table 1.

Another embodiment of the invention provides a method to aid in determining a ratio of alleles at a polymorphic locus in a sample. Any nucleic acid molecule, including genomic DNA, which comprises one or more polymorphic locus is used as a template in a single base extension reaction with an extension primer, forming a labeled extension primer. The extension primer comprises a 3' portion and a 5' portion. The 3' portion is complementary to the nucleic acid molecule and terminates one nucleotide 5' to the polymorphic locus. The 5' portion is not complementary to the nucleic acid molecule. A labeled dideoxynucleotide which is complementary to the polymorphic locus is coupled to the 3' end of the extension primer. Each type of dideoxynucleotide present in the reaction bears a distinct label. The 5' portion of the extension primer is

hybridized to one or more probes which are immobilized to known locations on a solid support.

These and other embodiments of the invention which are described in more detail below provide the art with methods and tools for rapidly and easily determining
5 genotypes of individuals and allele frequencies in populations.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a diagram of the universal array. The solid substrate (*e.g.*, a glass slide) is depicted on the left, and different oligonucleotide tags ("A", "B", "C", etc.) are shown attached to the solid substrate. The nucleotide sequence on the right-hand end of each
10 oligonucleotide tag ("Tag A", "Tag B", "Tag C") is arbitrary unique sequence; that is, it is designed and synthesized to be unique to each oligonucleotide tag.

Fig. 2 is a diagram depicting a locus-specific tagged oligonucleotide. The nucleotide sequence at the left-hand end is complementary to the arbitrary sequence of one of the oligonucleotide tags depicted in Fig. 1. The nucleotide sequence at the right-
15 hand end is complementary to the amplification product of a known polymorphic locus (*e.g.*, a single nucleotide polymorphism (SNP)). Therefore, locus-specific tagged oligonucleotide "A" comprises a nucleotide sequence complementary to the arbitrary sequence of the "Tag A" oligonucleotide tag depicted in Fig. 1, and also comprises sequence complementary to SNP "A".

20 Fig. 3 is a diagram showing the hybridization of the locus-specific tagged oligonucleotide to the amplification product. The locus-specific sequence (right hand end) of the oligonucleotide is designed so that it terminates one nucleotide immediately before (5' of) the nucleotide to be genotyped (shown in box).

Fig. 4 is a diagram depicting the labeling of the locus-specific tagged
25 oligonucleotide-amplification primer complex via single base extension. During the reaction, a single labeled ddNTP complementary to the queried nucleotide is enzymatically added to the 3' end of the locus-specific tagged oligonucleotide. The nucleotide is shown in the box.

Fig. 5 is a diagram depicting the hybridization of the complex of the amplification product and the locus-specific tagged oligonucleotide to the oligonucleotide tags on the array. The solid substrate to which the oligonucleotide tags of the array are bound is shown on the left, with the individual addresses labeled as "A", "B", etc. Each oligonucleotide tag is shown at its address. The locus-specific tagged oligonucleotide is shown hybridized to the oligonucleotide tag, and the amplification product is in turn bound to the locus-specific tagged oligonucleotide. The locus-specific tagged oligonucleotide is bound to a labeled (■, ●, etc.) nucleotide as a result of single base extension. Although a single complex is shown at each address, in reality, many such oligonucleotide tags are located at each address; that is, the substrate surface at address "A" has many copies of oligonucleotide tag "A" attached to it, etc.

Fig. 6 is a diagram depicting the hybridization as in Fig. 5, but the sample at address "B" is heterozygous for the queried nucleotide.

Fig. 7 is a schematic showing the combined use of amplification, single base extension of a tagged primer, and hybridization to a tag array.

Fig. 8 shows a quantitative measurement of allele frequency. Template-T (5'-TGCTGAATATTCAGATTCTCTAGTGCTACCTGAAAGATCCTG-3'; SEQ ID NO: 1) and Template-G (5'-TGCTGAATATTCAGATTCTCGAGTGCTACCTGAAAGATCCTG-3'; SEQ ID NO: 2) were mixed at different ratios (6 nM /60 nM, 6 nM /18 nM, 6 nM /6 nM, 18 nM /6 nM, 60 nM /6 nM, 180 nM /6 nM). Six SBE primers (5'-CACCATGCTCACAATGAATGCAGGATCTTTCAGGTAGCACT-3' (SEQ ID NO: 3); 5'-GATAATTCTCTGATAGGCCGCGAGGATCTTTCAGGTAGCACT-3' (SEQ ID NO: 4); 5'-GACTACGATGTGATCCGTGTCAGGATCTTTCAGGTAGCACT-3' (SEQ ID NO: 5);

5'-GAACGCAGTTATCAGACTCTCAGGATCTTTCAGGTAGCACT-3' (SEQ ID NO: 6);

5'-CGAGGACATGGAGTCACATCCAGGATCTTTCAGGTAGC-ACT-3' (SEQ ID NO: 7); and

- 5 5'-GCTAGGCATTCTCCAGTGTCTCAGGATCTTTCAGGTAGCACT-3' (SEQ ID NO: 8)) were separately added to six SBE reactions which contain the mixed templates of different ratios. The SBE primers were extended in the presence of biotin-labeled ddATP and fluorescein-labeled ddCTP (see Examples) and pooled and hybridized to the tag array. The intensity ratio of the two colors (the y-axis) were plotted against the ratio
10 of the mixed two templates (the x-axis).

Fig. 9 shows a clustering analysis of the tag array hybridization results in 44 individuals at marker GMP-140.25.

DETAILED DESCRIPTION OF THE INVENTION

- The invention features a generic or universal genotyping array, consisting of
15 oligonucleotide tags attached to a solid substrate (Fig. 1). Each address in the array (e.g., "A", "B", "C", etc.) has an oligonucleotide tag associated with it. The oligonucleotide tag at a given address is attached to the solid substrate, and comprises a unique arbitrary nucleotide sequence. That is, the nucleotide sequence is unique for the oligonucleotide tag at each address, i.e., the nucleotide sequence for "tag A" is different
20 from the nucleotide sequence for all other tags in the array. The nucleotide sequence for each tag is arbitrary in that it can be any sequence, provided that it is different from the nucleotide sequence for every other tag in the array. Preferably the oligonucleotide tag is from about 20 to about 50 nucleotides in length. It may also be desirable to design the nucleotide sequence of the oligonucleotide tag such that it does not facilitate an
25 undesirable interaction, e.g., with the target nucleic acid molecule (amplified product).

The oligonucleotide array is used in conjunction with locus-specific tagged oligonucleotides. Each oligonucleotide tag in the array corresponds to a locus-specific tagged oligonucleotide. One end (the 5' end) of the locus-specific tagged

oligonucleotide comprises a nucleotide sequence complementary to the unique arbitrary sequence of its corresponding oligonucleotide tag (Fig. 2). Preferably, this sequence is from about 20 to about 30 nucleotides long. The other end (the 3' end) of the locus-specific tagged oligonucleotide is complementary to a target nucleic acid molecule comprising a nucleotide to be queried, e.g., a polymorphic nucleotide. Preferably, the 3' end of locus-specific tagged oligonucleotide is synthesized such that when hybridized to the target nucleic acid molecule the locus-specific tagged oligonucleotide terminates one nucleotide 5' to the nucleotide to be queried. The portion of the locus-specific tagged oligonucleotide which hybridizes to the target nucleic acid molecule is preferably from about 15 to about 30 nucleotides long. For example, the 5' end of locus-specific tagged oligonucleotide "A" would be complementary to the unique arbitrary sequence at the end of the oligonucleotide tag "A" which is bound to address "A" in the array. The 3' end of locus-specific tagged oligonucleotide "A" would be complementary to the polynucleotide sequence 5' of the nucleotide to be queried in target "A".

To genotype a nucleic acid sample from an individual at locus "A", amplification primers specific for the region containing locus "A" are used to amplify the nucleic acid molecules in the sample. Locus-specific tagged oligonucleotides complementary to the nucleotide sequence 5' of locus "A" are combined with the amplification products under conditions suitable for hybridization (Fig. 3). The hybridization complex is subjected to single base extension. The four types of ddNTPs in the reaction mixture have different labels (*e.g.*, four different fluorescent tags, *e.g.*, the ddATPs would have an attached fluorophore that fluoresced at a first wavelength, the ddCTPs would have an attached fluorophore that fluoresced at a second wavelength, the ddGTPs would have an attached fluorophore that fluoresced at a third wavelength, and the ddTTPs would have an attached fluorophore that fluoresced at a fourth wavelength). During the single base extension reaction, a single ddNTP is attached (Fig. 4), resulting in the formation of a complex composed of the locus-specific tagged oligonucleotide extended with the labeled ddNTP and the amplification product.

After the single base extension reaction, the complex of the labeled (extended) locus-specific tagged oligonucleotide and the amplification product is hybridized to the array (Fig. 5). The oligonucleotide tag "A" at address "A" selectively hybridizes to its corresponding locus-specific tagged oligonucleotide (now extended with a labeled ddNTP), the oligonucleotide tag "B" at address "B" selectively hybridizes to its corresponding locus-specific tagged oligonucleotide (now extended with a labeled ddNTP), etc. The array is assayed to determine which label(s) is (are) present at which address on the array. For instance, if address "A" fluoresced at the same wavelength as the label on the ddATP, then the amplification product clearly contained a "T" at the queried nucleotide (because the single base extension reaction attaches the ddNTP complementary to the queried nucleotide). Fluorescence at a wavelength which is the same as the ddCTP label would indicate that the genotype was a "G", etc. Detection of two peaks within the wavelength emitted would indicate that different nucleotides were present at the queried position in the sample, e.g., that the individual was heterozygous at that locus.

An advantage of the array and method described herein is that many addresses can be assayed simultaneously, producing genotyping data for many different genetic loci, e.g., SNPs. By utilizing a predefined set of locus-specific tagged oligonucleotides, e.g., a set specific for assaying a set of genetic diseases, a single array can be utilized for a particular purpose, and by utilizing a different set of locus-specific tagged oligonucleotides which correspond to the same tags on the array, the same array can be utilized for a different purpose. The universal chip serves as the repository of a set of addresses to which the locus-specific tagged oligonucleotides (along with the labeled, genotyped SNPs) hybridize in a planned, predetermined manner. The array and set(s) of locus-specific tagged oligonucleotides can therefore be used as components in kits for the purposes of sequencing and genotyping. Sets of locus-specific tagged oligonucleotides can therefore be used in combination with arrays as described herein for use in forensics, identification of individuals, and disease diagnosis/prognosis.

The present invention provides a convenient and accurate way of determining the genotype of an individual at a polymorphic locus or the frequency of alleles in a population. One embodiment of the method involves three steps: (1) amplification of a polymorphic locus, (2) primer extension of a sequence-tagged primer with distinct
5 labels for different polynucleotides at the polymorphic locus, and (3) hybridization to a tag array. The amount of each distinct label can be determined at known positions of the tag array. Each tag represents a distinct polymorphic locus and each distinct label represents a distinct allelic form at the polymorphic locus. The method permits the simultaneous determination of a genotype at multiple loci, as well as the determination
10 of allele frequencies in a population. Another embodiment employs just steps 2 and 3.

Advantages of the disclosed method include that just one generic tag array can be used to genotype any genetic marker, *i.e.*, no specific customized genotyping chip is needed. In addition, the pre-selected probe sequences synthesized on the tag chip guarantee good hybridization results between the probe and the tag. Moreover, the two
15 color or multiple color approach used in this assay provides accurate measurement of the allele frequency in the samples tested. This means very reliable genotype results can be obtained not only for individual samples, but also for pooled samples.

A pair of primers or a single primer can be used to amplify a region of a nucleic acid in a sample. The sample may be from a single individual or may be from a
20 population of individuals. The region which is amplified includes a polymorphic locus. The step of amplification is not specific for a particular allele. However, the amplification is designed to specifically amplify regions of double stranded or single stranded nucleic acids which contain polymorphic loci.

The amplification step may be carried out using any technique known in the art.
25 One preferred technique is polymerase chain reaction (PCR) in which DNA is amplified logarithmically. As is known in the art, each primer of a pair of amplification primers hybridizes to, and is preferably complementary to, opposite strands of an allele. It is preferred that the primers hybridize to a double stranded nucleic acid in locations which are not more than 2 kb apart, and preferably which are much closer together, such as not

more than 1 kb, 0.5 kb, 0.2 kb, 0.1 kb, 0.01 kb or 0.001 kb apart. A suitable DNA polymerase can be used as is known in the art. Thermostable polymerases are particularly convenient for thermal cycling of rounds of primer hybridization, polymerization, and melting. Amplification of single stranded nucleic acids can also
5 be employed.

After the amplification it is desirable to remove and/or degrade any excess primers and nucleotides. This can be done by washing and/or enzymatic degradation, using such enzymes as endonuclease I and alkaline phosphatase, for example. Other techniques, such as chromatography, magnetic beads, and avidin- or streptavidin-
10 conjugated beads, as are known in the art for accomplishing the removal can also be used. It is not necessary to remove or destroy one of two strands of an amplified DNA product.

The primer extension step of the method is the one which provides allele-specificity to the method. The primer is designed to terminate one nucleotide 5' to the
15 polymorphic locus. The primer is hybridized to the denatured amplified double stranded DNA. When the primer is extended by a single base using dideoxynucleotides and a DNA polymerase, the dideoxynucleotide which is complementary to the nucleotide at the polymorphic locus is added. Again, any DNA-dependent DNA polymerase can be used. These include, but are not limited to, *E. coli* DNA polymerase
20 I, Klenow fragment of polymerase I, T4 DNA polymerase, T7 DNA polymerase, *T. aquaticus* DNA polymerase. This reaction is preferably performed at the T_M of the primer with the template to enhance product formation.

One configuration for carrying out the primer extension step utilizes two different primers which each hybridize to opposite strands of an amplified double
25 stranded DNA. Each primer terminates one nucleotide 5' to the polymorphic locus. The primer extension reaction may be more robust with one strand as a template than the other. In addition, the information obtained from the second strand should confirm the information obtained from the first strand.

An alternative method for primer extension involves use of reverse transcriptase and one or two primers which hybridize 3' to the polymorphic locus. This method may be desirable in cases where "forward" direction primer extension is less robust than is desirable.

- 5 Each different dideoxynucleotide present in the single base extension reaction is uniquely labeled. The unique label can be detected and its amount will be proportional to the amount of the particular allele containing the corresponding deoxynucleotide in the sample. If the sample is from a single individual, the nucleotide bases present at the polymorphic locus can be determined. If the sample is from a population of individuals
- 10 the allele frequency in the population can be determined.

- The ability to perform the method of the present invention in a multiplex manner for a number of different polymorphic loci simultaneously is due to the sequence tags which are present on the extension primers at their 5' ends. The sequence tags permit the method operator to ultimately sort the products of multiplex amplification and
- 15 multiplex primer base extension to different locations on an array. Each sequence tag on an extension primer is used only for a single polymorphic locus. Thus the products of primer extension reactions can be separately analyzed because they can be hybridized to distinct known locations on an array.

- The sequence tags are typically totally unrelated to the sequences of the
- 20 polymorphic alleles which are being analyzed. The sequence tags are chosen for their favorable hybridization characteristics. The tags are typically selected so that they have similar hybridization characteristics and minimal cross-hybridization to other tag sequences. Each sequence tag is attached to a specific gene or genetic marker, and then serves as a label for that particular gene or genetic marker. A generic tag array,
- 25 corresponding to the pre-selected tag sequences is fabricated and used to detect the presence or absence or ratio of specific allelic forms in a test sample. See application Serial No. 08/626,285 filed April 4, 1996, and EP application no. 97302313.8 which are expressly incorporated by reference herein.

The labels which are used can be any which are known in the art. These include radiolabels, fluorescent labels, enzyme labels, epitope labels, and high affinity binding partner labels. Examples include isotopically labeled nucleotides, fluorescein-labeled nucleotides, biotin-labeled nucleotides, digoxin labeled nucleotides. A different label is
5 assigned to each base dideoxynucleotide in the single base extension reaction. Two, three, or four different labels can be used in the reaction. The different labels can be all of the same type, *e.g.*, enzyme labels, or they can be mixed types.

Hybridization of the 5' portion of the extension primers (the tag sequences) to one or more probes which are immobilized to known locations on a solid support is also
10 contemplated. Hybridization can be performed under standard conditions known in the art for obtaining robust signals at high specificity. Standard washing conditions can also be employed. Detection of hybridization of the extension primers can be done using standard means, depending on the type of labels used. For example, fluorescence can be detected and quantified using optical detection means. Radiolabels can be
15 detected using autoradiography or scintillation counting. Enzyme labels can be detected using enzymatic reactions and assaying for the final product of the enzyme reaction. Antigenic labels can be used using immunological detection means. Affinity binding partners such as streptavidin or avidin and biotin can also be used as a label.

The reactions of the present invention can be performed in a single or multiplex
20 format. For example, the amplification step can be performed using up to 20, 30, 40, 50, 75, 100, 150, 200, 250, or 300 different primer pairs to amplify a corresponding number of polymorphic markers. These can be pooled for the single base extension reaction, if desired. Pooling for the hybridization step is desirable so that thousands of hybridizations can be done simultaneously.

25 In an alternative embodiment the amplification step can be omitted. Thus, if sufficient DNA is available, the single base extension reaction can be performed directly on genomic DNA. In another particular embodiment, amplification of the entire genome can be performed using random primers.

Sets of primers according to the present invention comprise an amplification pair and an extension primer. These are used together in a method for determining a ratio of nucleotides present at a polymorphic locus. These may be packaged in a single container, preferably a divided container or package. The pair of primers amplify a
5 region of double stranded DNA which comprises a polymorphic locus. The extension primer has two portions, a 3' portion which is complementary to a portion of the region of double stranded DNA which contains the polymorphic locus and a 5' portion which is not complementary to the region of double stranded DNA. The 5' region is the tag sequence which is complementary to the tag array which is used to sort and analyze the
10 products of the single base extension reaction. The 3' end of the single base extension primer terminates one nucleotide 5' to the polymorphic locus.

Kits according to the present invention may contain one or more sets of primers as described above. The kit may also contain a solid support comprising at least one probe which is attached to the solid support. The one or more probes are
15 complementary to the 5' portion of the extension primer, *i.e.*, to the tag sequences. Solid supports, according to the present invention include beads, microtiter plates, and arrays.

Hybridizing Nucleic Acids to Arrays of Allele-Specific Probes

“Hybridization” refers to the formation of a bimolecular complex of two
20 different nucleic acids through complementary base pairing. Complementary base pairing occurs through non-covalent bonding, usually hydrogen bonding, of bases that specifically recognize other bases, as in the bonding of complementary bases in double-stranded DNA. In this invention, hybridization is carried out between a target nucleic acid, which is prepared from the nucleic acid sample by allele-specific amplification,
25 and at least two probes which have been immobilized on a substrate to form an array.

One of skill in the art will appreciate that an enormous number of array designs are suitable for the practice of this invention. An array will typically include a number of probes that specifically hybridize to the sequences of interest (tags). In addition, it is

preferred that the array include one or more control probes. In one embodiment, the array is a high density array. A high density array is an array used to hybridize with a target nucleic acid sample to detect the presence of a large number of allelic markers, preferably more than 10, more preferably more than 100, and most preferably more than 1000 allelic markers.

High density arrays are suitable for quantifying small variations in the frequency of an allelic marker in the presence of a large population of heterogeneous nucleic acids. Such high density arrays can be fabricated either by *de novo* synthesis on a substrate or by spotting or transporting nucleic acid sequences onto specific locations of a substrate. Both of these methods produce nucleic acids which are immobilized on the array at particular locations. Nucleic acids can be purified and/or isolated from biological materials, such as a bacterial plasmid containing a cloned segment of a sequence of interest. Suitable nucleic acids can also be produced by amplification of templates or by synthesis. As a nonlimiting illustration, polymerase chain reaction and/or *in vitro* transcription, are suitable nucleic acid amplification methods.

The term "target nucleic acid" refers to a nucleic acid (either synthetic or derived from a biological sample or nucleic acid sample), to which the probe is designed to specifically hybridize. In this invention, such target nucleic acids are the same as the sequence tags. It is either the presence or absence of the target nucleic acid that is to be detected, or the amount of the target nucleic acid that is to be quantified. The target nucleic acid has a sequence that is complementary to the nucleic acid sequence of the corresponding probe directed to the target. The term "target nucleic acid" can refer to the specific subsequence of a larger nucleic acid to which the probe is directed or to the overall sequence (*e.g.*, gene or mRNA) whose presence it is desired to detect. The difference in usage will be apparent from context.

As used herein a "probe" is defined as a nucleic acid, capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe can include natural (*i.e.* A, G, U, C, or T) or

modified bases (*e.g.*, 7-deazaguanosine, inosine, etc.). A probe can also include an oligonucleotide. An oligonucleotide is a single-stranded nucleic acid of 2 to *n* bases, where *n* can be any integer less than 1000. Nucleic acids can be cloned or synthesized using any technique known in the art. They can also include non-naturally occurring nucleotide analogs, such as those which are modified to improve hybridization, and peptide nucleic acids. In addition, the bases in probes may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages.

10 Probe Design

An array includes "test probes", also termed "oligonucleotide tags" herein. Test probes can be oligonucleotides that range from about 5 to about 45 or 5 to about 500 nucleotides, more preferably from about 10 to about 40 nucleotides and most preferably from about 15 to about 40 nucleotides in length. In other particularly preferred embodiments the probes are 20 to 25 nucleotides in length. In another embodiment, test probes are double or single stranded DNA sequences. DNA sequences can be isolated or cloned from natural sources or amplified from natural sources using natural nucleic acids as templates. However, *in situ* synthesis of probes on the arrays is preferred. The probes have sequences complementary to particular subsequences of the genes whose allelic markers they are designed to detect. Thus, the test probes are capable of specifically hybridizing to the target nucleic acid they are designed to detect.

The term "perfect match probe" refers to a probe which has a sequence designed to be perfectly complementary to a particular target sequence. The probe is typically perfectly complementary to a portion (subsequence) of the target sequence. The perfect match probe can be a "test probe," a "normalization control probe," an expression level control probe and the like. A perfect match control or perfect match probe is, however, distinguished from a "mismatch control" or "mismatch probe" or "mismatch control probe."

In addition to test probes that bind the target nucleic acid(s) of interest, the high density array can contain a number of control probes. The control probes fall into two categories: normalization controls and mismatch controls.

Normalization controls are oligonucleotide or other nucleic acid probes that are complementary to labeled reference oligonucleotides or other nucleic acid sequences that are added to the nucleic acid sample. The signals obtained from the normalization controls after hybridization provide a control for variations in hybridization conditions, label intensity, "reading" efficiency, and other factors that may cause the signal of a perfect hybridization to vary between arrays. In a preferred embodiment, signals (*e.g.*, fluorescence intensity) read from all other probes in the array are divided by the signal (*e.g.*, fluorescence intensity) from the control probes, thereby normalizing the measurements.

Virtually any probe can serve as a normalization control. However, it is recognized that hybridization efficiency varies with base composition and probe length. Preferred normalization probes are selected to reflect the average length of the other probes present in the array; however, they can be selected to cover a range of lengths. The normalization control(s) can also be selected to reflect the (average) base composition of the other probes in the array; however in a preferred embodiment, only one or a few normalization probes are used and they are selected such that they hybridize well (*i.e.* no secondary structure) and do not match any target-specific probes.

Mismatch controls can also be provided for the probes to the target alleles or for normalization controls. The terms "mismatch control" or "mismatch probe" or "mismatch control probe" refer to a probe whose sequence is deliberately selected not to be perfectly complementary to a particular target sequence. Mismatch controls are oligonucleotide probes or other nucleic acid probes identical to their corresponding test or control probes except for the presence of one or more mismatched bases. A mismatched base is a base selected so that it is not complementary to the corresponding base in the target sequence to which the probe would otherwise specifically hybridize. One or more mismatches are selected such that under appropriate hybridization

conditions (*e.g.*, stringent conditions) the test or control probe would be expected to hybridize with its target sequence, but the mismatch probe would not hybridize (or would hybridize to a significantly lesser extent). Preferred mismatch probes contain a central mismatch. Thus, for example, where a probe is a 20 mer, a corresponding
 5 mismatch probe will have the identical sequence except for a single base mismatch (*e.g.*, substituting a G, a C, or a T for an A) at any of positions 6 through 14 (the central mismatch).

For each mismatch control in a high-density array there typically exists a corresponding perfect match probe that is perfectly complementary to the same
 10 particular target sequence. The mismatch may comprise one or more bases. While the mismatch(s) may be located anywhere in the mismatch probe, terminal mismatches are less desirable, as a terminal mismatch is less likely to prevent hybridization of the target sequence. In a particularly preferred embodiment, the mismatch is located at or near the center of the probe such that the mismatch is most likely to destabilize the duplex with
 15 the target sequence under the test hybridization conditions.

Mismatch probes provide a control for non-specific binding or cross-hybridization to a nucleic acid in the sample other than the target to which the probe is directed. Mismatch probes thus indicate whether or not a hybridization is specific. For example, if the target is present, the perfect match probes should be consistently
 20 brighter than the mismatch probes. The difference in intensity between the perfect match and the mismatch probe ($I_{(PM)} - I_{(MM)}$) provides a good measure of the concentration of the hybridized material.

The array can also include sample preparation/amplification control probes. These are probes that are complementary to subsequences of control genes selected
 25 because they do not normally occur in the nucleic acids of the particular biological sample being assayed. Suitable sample preparation/amplification control probes include, for example, probes to bacterial genes (*e.g.*, Bio B) where the sample in question is from a eukaryote.

In a preferred embodiment, oligonucleotide probes in the high density array are selected to bind specifically to the nucleic acid target to which they are directed with minimal non-specific binding or cross-hybridization under the particular hybridization conditions utilized. Because the high density arrays of this invention can contain in
5 excess of 100,000 or even 1,000,000 different probes, it is possible to provide every probe of a characteristic length that binds to a particular nucleic acid sequence.

Forming High Density Arrays

High density arrays are particularly useful for monitoring the presence of allelic markers. The fabrication and application of high density arrays in gene expression
10 monitoring have been disclosed previously in, for example, WO 97/10365, WO 92/10588, U.S. Application Ser. No. 08/772,376 filed December 23, 1996; serial number 08/529,115 filed on September 15, 1995; serial number 08/168,904 filed December 15, 1993; serial number 07/624,114 filed on December 6, 1990, serial number 07/362,901 filed June 7, 1990, and in U.S. 5,677,195, all incorporated herein for
15 all purposes by reference. In some embodiments using high density arrays, high density oligonucleotide arrays are synthesized using methods such as the Very Large Scale Immobilized Polymer Synthesis (VLSIPS) disclosed in U.S. Pat. No. 5,445,934 incorporated herein for all purposes by reference. Each oligonucleotide occupies a known location on a substrate. A nucleic acid target sample is hybridized with a high
20 density array of oligonucleotides and then the amount of target nucleic acids hybridized to each probe in the array is quantified.

Synthesized oligonucleotide arrays are particularly preferred for this invention. Oligonucleotide arrays have numerous advantages over other methods, such as efficiency of production, reduced intra- and inter array variability, increased information
25 content, and high signal-to-noise ratio.

Preferred high density arrays comprise greater than about 100, preferably greater than about 1000, more preferably greater than about 16,000, and most preferably greater than 65,000 or 250,000 or even greater than about 1,000,000 different oligonucleotide

probes, preferably in less than 1 cm² of surface area. The oligonucleotide probes range from about 5 to about 50 or about 500 nucleotides, more preferably from about 10 to about 40 nucleotides, and most preferably from about 15 to about 40 nucleotides in length.

- 5 Methods of forming high density arrays of oligonucleotides, peptides and other polymer sequences with a minimal number of synthetic steps are known. The oligonucleotide analogue array can be synthesized on a solid substrate by a variety of methods, including, but not limited to, light-directed chemical coupling and mechanically directed coupling. See Pirrung *et al.*, U.S. Patent No. 5,143,854 (see also
- 10 PCT Application No. WO 90/15070) and Fodor *et al.*, PCT Publication Nos. WO 92/10092 and WO 93/09668 and U.S. Ser. No. 07/980,523, which disclose methods of forming vast arrays of peptides, oligonucleotides and other molecules using, for example, light-directed synthesis techniques. See also, Fodor *et al.*, *Science*, 251, 767-77 (1991). These procedures for synthesis of polymer arrays are now referred to as
- 15 VLSIPS™ procedures. Using the VLSIPS™ approach, one heterogeneous array of polymers is converted, through simultaneous coupling at a number of reaction sites, into a different heterogeneous array. See, U.S. Application Serial Nos. 07/796,243 and 07/980,523.

- The development of VLSIPS™ technology as described in the above-noted U.S.
- 20 Patent No. 5,143,854 and PCT patent publication Nos. WO 90/15070 and 92/10092, is considered pioneering technology in the fields of combinatorial synthesis and screening of combinatorial libraries. More recently, patent application Serial No. 08/082,937, filed June 25, 1993, describes methods for making arrays of oligonucleotide probes that can be used to check or determine a partial or complete sequence of a target nucleic acid
- 25 and to detect the presence of a nucleic acid containing a specific oligonucleotide sequence.

In brief, the light-directed combinatorial synthesis of oligonucleotide arrays on a glass surface proceeds using automated phosphoramidite chemistry and chip masking techniques. In one specific implementation, a glass surface is derivatized with a silane

reagent containing a functional group, *e.g.*, a hydroxyl or amine group blocked by a photolabile protecting group. Photolysis through a photolithographic mask is used selectively to expose functional groups which are then ready to react with incoming 5'-photoprotected nucleoside phosphoramidites. The phosphoramidites react only with those sites which are illuminated (and thus exposed by removal of the photolabile blocking group). Thus, the phosphoramidites only add to those areas selectively exposed from the preceding step. These steps are repeated until the desired array of sequences have been synthesized on the solid surface. Combinatorial synthesis of different oligonucleotide analogues at different locations on the array is determined by the pattern of illumination during synthesis and the order of addition of coupling reagents.

In the event that an oligonucleotide analogue with a polyamide backbone is used in the VLSIPS™ procedure, it is generally inappropriate to use phosphoramidite chemistry to perform the synthetic steps, since the monomers do not attach to one another via a phosphate linkage. Instead, peptide synthetic methods are substituted. See, *e.g.*, Pirrung *et al.* U.S. Pat. No. 5,143,854.

Peptide nucleic acids are commercially available from, *e.g.*, Biosearch, Inc. (Bedford, MA) which comprise a polyamide backbone and the bases found in naturally occurring nucleosides. Peptide nucleic acids are capable of binding to nucleic acids with high specificity, and are considered "oligonucleotide analogues" for purposes of this disclosure.

Additional methods which can be used to generate an array of oligonucleotides on a single substrate are described in co-pending Applications Ser. No. 07/980,523, filed November 20, 1992, and 07/796,243, filed November 22, 1991 and in PCT Publication No. WO 93/09668. In the methods disclosed in these applications, reagents are delivered to the substrate by either (1) flowing within a channel defined on predefined regions or (2) "spotting" on predefined regions or (3) through the use of photoresist. However, other approaches, as well as combinations of spotting and flowing, can be employed. In each instance, certain activated regions of the substrate

are mechanically separated from other regions when the monomer solutions are delivered to the various reaction sites.

A typical "flow channel" method applied to the compounds and libraries of the present invention can generally be described as follows. Diverse polymer sequences are synthesized at selected regions of a substrate or solid support by forming flow channels on a surface of the substrate through which appropriate reagents flow or in which appropriate reagents are placed. For example, assume a monomer "A" is to be bound to the substrate in a first group of selected regions. If necessary, all or part of the surface of the substrate in all or a part of the selected regions is activated for binding by, for example, flowing appropriate reagents through all or some of the channels, or by washing the entire substrate with appropriate reagents. After placement of a channel block on the surface of the substrate, a reagent having the monomer A flows through or is placed in all or some of the channel(s). The channels provide fluid contact to the first selected regions, thereby binding the monomer A on the substrate directly or indirectly (via a spacer) in the first selected regions.

Thereafter, a monomer "B" is coupled to second selected regions, some of which can be included among the first selected regions. The second selected regions will be in fluid contact with a second flow channel(s) through translation, rotation, or replacement of the channel block on the surface of the substrate; through opening or closing a selected valve; or through deposition of a layer of chemical or photoresist. If necessary, a step is performed for activating at least the second regions. Thereafter, the monomer B is flowed through or placed in the second flow channel(s), binding monomer B at the second selected locations. In this particular example, the resulting sequences bound to the substrate at this stage of processing will be, for example, A, B, and AB. The process is repeated to form a vast array of sequences of desired length at known locations on the substrate.

After the substrate is activated, monomer A can be flowed through some of the channels, monomer B can be flowed through other channels, a monomer C can be flowed through still other channels, *etc.* In this manner, many or all of the reaction

regions are reacted with a monomer before the channel block must be moved or the substrate must be washed and/or reactivated. By making use of many or all of the available reaction regions simultaneously, the number of washing and activation steps can be minimized.

- 5 One of skill in the art will recognize that there are alternative methods of forming channels or otherwise protecting a portion of the surface of the substrate. For example, according to some embodiments, a protective coating such as a hydrophilic or hydrophobic coating (depending upon the nature of the solvent) is utilized over portions of the substrate to be protected, sometimes in combination with materials that facilitate
- 10 wetting by the reactant solution in other regions. In this manner, the flowing solutions are further prevented from passing outside of their designated flow paths.

- High density nucleic acid arrays can be fabricated by depositing presynthesized or natural nucleic acids in predetermined positions. Synthesized or natural nucleic acids are deposited on specific locations of a substrate by light directed targeting and
- 15 oligonucleotide directed targeting. Nucleic acids can also be directed to specific locations in much the same manner as the flow channel methods. For example, a nucleic acid A can be delivered to and coupled with a first group of reaction regions which have been appropriately activated. Thereafter, a nucleic acid B can be delivered to and reacted with a second group of activated reaction regions. Nucleic acids are
- 20 deposited in selected regions. Another embodiment uses a dispenser that moves from region to region to deposit nucleic acids in specific spots. Typical dispensers include a micropipette or capillary pin to deliver nucleic acid to the substrate and a robotic system to control the position of the micropipette with respect to the substrate. In other embodiments, the dispenser includes a series of tubes, a manifold, an array of pipettes
- 25 or capillary pins, or the like so that various reagents can be delivered to the reaction regions simultaneously.

Hybridization Conditions

The term "stringent conditions" refers to conditions under which a probe will hybridize to its target subsequence, but with only insubstantial hybridization to other sequences or to other sequences such that the difference may be identified. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

The T_m is the temperature, under defined ionic strength, pH, and nucleic acid concentration, at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. As the target sequences are generally present in excess, at T_m , 50% of the probes are occupied at equilibrium). Typically, stringent conditions will be those in which the salt concentration is at least about 0.01 to 1.0 M concentration of a Na or other salt at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides). Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide.

The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule substantially to or only to a particular nucleotide sequence or sequences under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) of DNA or RNA. It is generally recognized that nucleic acids are denatured by increasing the temperature or decreasing the salt concentration of the buffer containing the nucleic acids. Under low stringency conditions (*e.g.*, low temperature and/or high salt) hybrid duplexes (*e.g.*, DNA:DNA, RNA:RNA, or RNA:DNA) will form even where the annealed sequences are not perfectly complementary. Thus, specificity of hybridization is reduced at lower stringency. Conversely, at higher stringency (*e.g.*, higher temperature or lower salt) successful hybridization requires fewer mismatches.

One of skill in the art will appreciate that hybridization conditions can be selected to provide any degree of stringency. In a preferred embodiment, hybridization

is performed at low stringency, in this case in 6X SSPE-T at 37°C (0.005% Triton X-100), to ensure hybridization, and then subsequent washes are performed at higher stringency (*e.g.*, 1 X SSPE-T at 37°C) to eliminate mismatched hybrid duplexes.

Successive washes can be performed at increasingly higher stringency (*e.g.*, down to as low as 0.25 X SSPE-T at 37°C to 50°C) until a desired level of hybridization specificity is obtained. Stringency can also be increased by addition of agents such as formamide. Hybridization specificity can be evaluated by comparison of hybridization to the test probes with hybridization to the various controls that can be present (*e.g.*, expression level control, normalization control, mismatch controls, *etc.*).

10 In general, there is a tradeoff between hybridization specificity (stringency) and signal intensity. Thus, in a preferred embodiment, the wash is performed at the highest stringency that produces consistent results and that provides a signal intensity greater than approximately 10% of the background intensity. Thus, in a preferred embodiment, the hybridized array can be washed at successively higher stringency solutions and read
15 between each wash. Analysis of the data sets thus produced will reveal a wash stringency above which the hybridization pattern is not appreciably altered and which provides adequate signal for the particular oligonucleotide probes of interest.

The stability of duplexes formed between RNAs or DNAs are generally in the order of RNA:RNA > RNA:DNA > DNA:DNA, in solution. Long probes have better
20 duplex stability with a target, but poorer mismatch discrimination than shorter probes (mismatch discrimination refers to the measured hybridization signal ratio between a perfect match probe and a single base mismatch probe). Shorter probes (*e.g.*, 8-mers) discriminate mismatches very well, but the overall duplex stability is low.

Altering the thermal stability (T_m) of the duplex formed between the target and
25 the probe using, *e.g.*, known oligonucleotide analogues allows for optimization of duplex stability and mismatch discrimination. One useful aspect of altering the T_m arises from the fact that adenine-thymine (A-T) duplexes have a lower T_m than guanine-cytosine (G-C) duplexes, due in part to the fact that the A-T duplexes have two hydrogen bonds per base-pair, while the G-C duplexes have three hydrogen bonds per

base pair. In heterogeneous oligonucleotide arrays in which there is a non-uniform distribution of bases, it is not generally possible to optimize hybridization for each oligonucleotide probe simultaneously. Thus, in some embodiments, it is desirable to selectively destabilize G-C duplexes and/or to increase the stability of A-T duplexes.

- 5 This can be accomplished, *e.g.*, by substituting guanine residues in the probes of an array which form G-C duplexes with hypoxanthine, or by substituting adenine residues in probes which form A-T duplexes with 2,6 diaminopurine or by using tetramethyl ammonium chloride (TMACl) in place of NaCl.

- 10 Altered duplex stability conferred by using oligonucleotide analogue probes can be ascertained by following, *e.g.*, fluorescence signal intensity of oligonucleotide analogue arrays hybridized with a target oligonucleotide over time. The data allow optimization of specific hybridization conditions at, *e.g.*, room temperature.

- 15 Another way of verifying altered duplex stability is by following the signal intensity generated upon hybridization with time. Previous experiments using DNA targets and DNA chips have shown that signal intensity increases with time, and that the more stable duplexes generate higher signal intensities faster than less stable duplexes. The signals reach a plateau or "saturate" after a certain amount of time due to all of the binding sites becoming occupied. These data allow for optimization of hybridization, and determination of the best conditions at a specified temperature.

- 20 Methods of optimizing hybridization conditions are well known to those of skill in the art (*see, e.g., Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 24: Hybridization With Nucleic Acid Probes*, P. Tijssen, ed. Elsevier, N.Y., (1993)).

Signal Detection

- 25 The hybridized nucleic acids can be detected by detecting one or more labels attached to the target nucleic acids. The labels can be incorporated by any of a number of means well known to those of skill in the art. However, in a preferred embodiment, the label is incorporated by labeling the primers prior to the amplification step in the

preparation of the target nucleic acids. Thus, for example, polymerase chain reaction with labeled primers will provide a labeled amplification product.

Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical, or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (*e.g.*, DynabeadsTM), fluorescent dyes (*e.g.*, fluorescein, texas red, rhodamine, green fluorescent protein, and the like), radiolabels (*e.g.*, ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P), enzymes (*e.g.*, horseradish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.*, polystyrene, polypropylene, latex, etc.) beads. Patents teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

Means of detecting such labels are well known to those of skill in the art. Thus, for example, radiolabels can be detected using photographic film or scintillation counters, fluorescent markers can be detected using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting the reaction product produced by the action of the enzyme on the substrate, and colorimetric labels are detected by simply visualizing the colored label. One method uses colloidal gold label that can be detected by measuring scattered light.

Means of detecting labeled target nucleic acids hybridized to the probes of the array are known to those of skill in the art. Thus, for example, where a colorimetric label is used, simple visualization of the label is sufficient. Where a radioactive labeled probe is used, detection of the radiation (*e.g.* with photographic film or a solid state detector) is sufficient.

Detection of target nucleic acids which are labeled with a fluorescent label (*i.e.*, a "color tag") can be accomplished with fluorescence microscopy. The hybridized array can be excited with a light source at the excitation wavelength of the particular fluorescent label and the resulting fluorescence at the emission wavelength is detected.

The excitation light source can be a laser appropriate for the excitation of the fluorescent label.

The confocal microscope can be automated with a computer-controlled stage to automatically scan the entire high density array, *i.e.*, to sequentially examine individual probes or adjacent groups of probes in a systematic manner until all probes have been examined. Similarly, the microscope can be equipped with a phototransducer (*e.g.*, a photomultiplier, a solid state array, a CCD camera, *etc.*) attached to an automated data acquisition system to automatically record the fluorescence signal produced by hybridization to each oligonucleotide probe on the array. Such automated systems are described at length in U.S. Patent No: 5,143,854, PCT Application 20 92/10092, and copending U.S. Application Ser. No. 08/195,889, filed on February 10, 1994. Use of laser illumination in conjunction with automated confocal microscopy for signal detection permits detection at a resolution of better than about 100 μm , more preferably better than about 50 μm , and most preferably better than about 25 μm .

Two different fluorescent labels can be used in order to distinguish two alleles at each marker examined. In such a case, the array can be scanned two times. During the first scan, the excitation and emission wavelengths are set as required to detect one of the two fluorescent labels. For the second scan, the excitation and emission wavelengths are set as required to detect the second fluorescent label. When the results from both scans are compared, the genotype identification or allele frequency can be determined.

Quantification and Determination of Genotypes

The term "quantifying" when used in the context of quantifying hybridization of a nucleic acid sequence or subsequence can refer to absolute or to relative quantification. Absolute quantification can be accomplished by inclusion of known concentration(s) of one or more target nucleic acids (*e.g.*, control nucleic acids such as Bio B, or known amounts the target nucleic acids themselves) and referencing the hybridization intensity of unknowns with the known target nucleic acids (*e.g.*, through

generation of a standard curve). Alternatively, relative quantification can be accomplished by comparison of hybridization signals between two or more genes, or between two or more treatments to quantify the changes in hybridization intensity and, by implication, the frequency of an allele. Relative quantification can also be used to
5 merely detect the presence or absence of an allele in the target nucleic acids. In one embodiment, for example, the presence or absence of the two alleles of a marker can be determined by comparing the quantities of the first and second color tag at the known locations in the array, *i.e.*, on the solid support, which correspond to the allele-specific probes for the two alleles.

10 A preferred quantifying method is to use a confocal microscope and fluorescent labels. The GeneChip[®] system (Affymetrix, Santa Clara, CA) is particularly suitable for quantifying the hybridization; however, it will be apparent to those of skill in the art that any similar system or other effectively equivalent detection method can also be used.

15 Methods for evaluating the hybridization results vary with the nature of the specific probes used, as well as the controls. Simple quantification of the fluorescence intensity for each probe can be determined. This can be accomplished simply by measuring signal strength at each location (representing a different probe) on the high density array (*e.g.*, where the label is a fluorescent label, detection of the fluorescence
20 intensity produced by a fixed excitation illumination at each location on the array).

One of skill in the art, however, will appreciate that hybridization signals will vary in strength with efficiency of hybridization, the amount of label on the sample nucleic acid and the amount of the particular nucleic acid in the sample. Typically nucleic acids present at very low levels (*e.g.*, < 1 pM) will show a very weak signal. At
25 some low level of concentration, the signal becomes virtually indistinguishable from background. In evaluating the hybridization data, a threshold intensity value can be selected below which a signal is counted as being essentially indistinguishable from background.

The terms "background" or "background signal intensity" refer to hybridization signals resulting from non-specific binding, or other interactions, between the labeled target nucleic acids and components of the oligonucleotide array (*e.g.*, the oligonucleotide probes, control probes, the array substrate, *etc.*). Background signals may also be produced by intrinsic fluorescence of the array components themselves. A single background signal can be calculated for the entire array, or a different background signal may be calculated for each target nucleic acid. In a preferred embodiment, background is calculated as the average hybridization signal intensity for the lowest 5% to 10% of the probes in the array, or, where a different background signal is calculated for each target allele, for the lowest 5% to 10% of the probes for each allele. However, where the probes to a particular allele hybridize well and thus appear to be specifically binding to a target sequence, they should not be used in a background signal calculation. Alternatively, background may be calculated as the average hybridization signal intensity produced by hybridization to probes that are not complementary to any sequence found in the sample (*e.g.*, probes directed to nucleic acids of the opposite sense or to genes not found in the sample, such as bacterial genes where the sample is mammalian nucleic acids). Background can also be calculated as the average signal intensity produced by regions of the array that lack any probes at all. In a preferred embodiment, background signal is reduced by the use of a detergent (*e.g.*, C-TAB) or a blocking reagent (*e.g.*, sperm DNA, cot-1 DNA, *etc.*) during the hybridization to reduce non-specific binding. In a particularly preferred embodiment, the hybridization is performed in the presence of about 0.5 mg/ml DNA (*e.g.*, herring sperm DNA). The use of blocking agents in hybridization is well known to those of skill in the art (*see, e.g.*, Chapter 8 in P. Tijssen, *supra*).

The high density array can include mismatch controls. In a preferred embodiment, there is a mismatch control having a central mismatch for every probe in the array, except the normalization controls. It is expected that after washing in stringent conditions, where a perfect match would be expected to hybridize to the probe, but not to the mismatch, the signal from the mismatch controls should only reflect non-

specific binding or the presence in the sample of a nucleic acid that hybridizes with the mismatch. Where both the probe in question and its corresponding mismatch control show high signals, or the mismatch shows a higher signal than its corresponding test probe, there is a problem with the hybridization and the signal from those probes is ignored. For a given marker, the difference in hybridization signal intensity ($I_{\text{allele1}} - I_{\text{allele2}}$) between an allele-specific probe (perfect match probe) for a first allele and the corresponding probe for a second allele (or other mismatch control probe) is a measure of the presence of or concentration of the first allele. Thus, in a preferred embodiment, the signal of the mismatch probe is subtracted from the signal for its corresponding test probe to provide a measure of the signal due to specific binding of the test probe.

The concentration of a particular sequence can then be determined by measuring the signal intensity of each of the probes that bind specifically to that gene and normalizing to the normalization controls. Where the signal from the probes is greater than the mismatch, the mismatch is subtracted. Where the mismatch intensity is equal to or greater than its corresponding test probe, the signal is ignored (*i.e.*, the signal cannot be evaluated).

For each marker analyzed, the genotype can be unambiguously determined by comparing the hybridization patterns obtained for each of the two labels, *e.g.*, color tags employed (Fig. 8). If hybridization is indicated for one color tag to its corresponding allele-specific probe (*e.g.*, "A") but not for the other color tag (*e.g.*, "G") (pattern at left in Fig. 8), then the indicated genotype of a diploid organism would be homozygous A/A. If hybridization is indicated only for the other color tag to its corresponding allele-specific probe (*e.g.*, "G") (pattern at center in Fig. 8), then the indicated genotype of a diploid organism would be homozygous G/G. If hybridization is indicated for both color tags to their corresponding allele-specific probes (pattern at right in Fig. 8), then the indicated genotype of a diploid organism would be heterozygous (A/G).

Marginal detection of hybridization, indicated by an intermediate positive result (*e.g.*, less than 1%, or from 1-5%, or from 1-10%, or from 2-10%, or from 5-10%, or from 1-20%, or from 2-20%, or from 5-20%, or from 10-20% of the average of all

positive hybridization results obtained for the entire array) may indicate either cross-hybridization or cross-amplification, depending on the overall hybridization pattern as indicated in Fig. 8. However, these can be distinguished by the unique pattern observed. Further procedures for data analysis are disclosed in U.S. Application 08/772,376, previously incorporated for all purposes.

HuSNP and other marker-specific arrays have been designed and used in genetic studies⁹⁻¹⁰. But the method developed in this study provides several advantages in dealing with many different genetic applications: (1) arrays based on a single generic design can be used to genotype different sets of genetic markers because no specific customized genotyping array is needed; (2) the pre-selected probe sequences synthesized on the tag array help ensure good hybridization results; (3) accurate quantitative measurement of the allele frequency in the tested samples can be achieved. Thus, reliable genotype results can be obtained not only for individual samples, but also for pooled samples. Besides SBE, other assays can be coupled with tag array assay, for example, oligonucleotide ligation assay (OLA)¹⁹⁻²¹, invasive cleavage of oligonucleotide probes assay²², allele specific PCR²³⁻²⁴.

Our current tag chip contains over 32,000 unique tag probes. For most of the genetic application, for example, detecting mutations in one particular gene, it doesn't need such high-density chip. Therefore, smaller chips with fewer tags on the chip are sought after. Alternatively, multiple tags corresponding to one particular marker can be designed as to build the redundancy to the assay to assure accurate genotyping. Or multiple sets of tags for one set of SNPs can be designed, thus multiple samples can be processed and analyzed with one chip. Our current assay uses a two-color labeling scheme. But a four-color labeling/scanning system should warrant the assay can be done in a single tube reaction.

For broader genetic applications, for example, a study needs to genotype 100s to 1000s genetic markers, amplifying the genetic loci with multiplexing PCR is still the best strategy. However, to genotype 1000s to 10,000s markers, pre-amplification of the interested genetic loci will be very labor-intensive and costly. A whole-genome

approach should be explored, for example, strategies involved using total human genomic DNA directly, or genomic DNA amplified using some general amplification methods, *e.g.*, primer-extension preamplification, PEP²⁵, or total cDNA. In fact, we have tried to use total human genomic DNA directly as the SBE template in our tag
 5 array assay. 24 out of the 38 of the markers that we tested gave good signals (data not shown). Nevertheless, some work is needed to solve both the sensitivity (signal intensity) and specificity (mis-priming) problems before the whole-genome approach becomes really useful.

The invention will be further illustrated by the following non-limiting examples.

10 The content of references cited herein is incorporated herein by reference in its entirety.

EXEMPLIFICATION

METHODS

Collection and Isolation of DNA From Samples

DNA samples were collected by GenNet as part of the ongoing Family Blood
 15 Pressure Program. Samples were collected with consent and IRB approval in both Tecumseh, MI and Loyola, IL FAMILIES. Ascertainment was based on identification of a proband in the top 15th (Tecumseh) or 20th (Loyola) percentile of the community's blood pressure distribution. Full phenotypic information was obtained for each individual. DNA was extracted from 5-10 ml of whole blood taken from each individual
 20 using the standard "salting-out" method (Gentra Systems).

Primer Design

For each SNP, primary PCR amplification primers were designed as described previously⁹. The SBE primer was designed in a manner that its 3' terminates one base before the polymorphic site. Primer 3.0 software package
 25 (<http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi>) was modified and used to pick SBE primers with batch sequences, at a predicted length of 20 (ranging from 18 to 26) nucleotide and melting temperature of 60°C (ranging from 54°C to 64°C). The SBE

primers were always picked from the forward direction first (i.e. 5' to the polymorphic site). If the SBE primer can't be picked from the forward direction, reverse direction is tried.

Multiplexing PCR

- 5 Specific genomic regions containing the 144 SNPs were amplified with 9 multiplex PCR reactions, each contains 50 ng of human genomic DNA, 0.1 μ M of each primer, 1 mM deoxynucleotide triphosphates (dNTPs), 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 5 mM $MgCl_2$ and 2 units of AmpliTaq Gold (Perkin Elmer) in a total value of 25 μ l. PCR was performed on a Thermo Cycler (MJ Research), with initial denaturation
- 10 of the DNA templates and Taq enzyme activation at 96°C for 10 minutes; followed by 40 cycles of denaturation at 94°C for 30 seconds, 57°C for 40 seconds, and 72°C for 1 minute and 30 seconds; and the final extension at 72°C for 10 minutes.

SBE Template Preparation

- 1 1 μ l of Exonuclease I (Amersham Life Science, 10 U/ μ l) and 1 μ l of Shrimp
- 15 Alkaline Phosphatase (Amersham Life Science, 1 U/ μ l) were added to a 25 μ l PCR products (see above), and incubated at 37°C for 1 hour. The enzyme activities were inactivated at 100°C for 15 minutes. The enzymatically treated samples were applied to a S-300 column (Pharmacia), as to further reduce the residual PCR primers and dNTPs, and replace the buffer with ddH₂O.

20 Multiplexing SBE Reaction

- SBE is carried out in a 33 μ l reaction, using 6 μ l of the template (see above), 1.5 nM of each SBE primer, 2.5 units of Thermo sequenase (Amersham), 52 mM Tris-HCl (pH 9.5), 6.5 mM $MgCl_2$, 25 μ M of fluorescein-N6-ddNTPs (NEN), 7.5 μ M biotin-N6-ddUTP or biotion-N6-dCTP, or 3.75 μ M biotin-N6-ddATP, and 10 μ M the
- 25 other cold ddNTPs.

Extension reaction was carried out on a Thermo Cycler (MJ Research), with 1 cycle of 96°C for 3 minutes, then 45 cycles of 94°C for 20 seconds and 58°C for 11 seconds.

After SBE reaction, 9 reactions from each sample were combined and mixed
 5 with 30 µl of 100 µg/ml glycogen (Boehringer Mannheim), 18.75 µl of 8 M LiCl (Sigma), and 1125 µl of pre-chilled (-20°C) ethanol (Abs.), and precipitated by centrifugation at the top speed (Eppendorf centrifuge 5415C) for 15 minutes at room temperature; precipitated samples were dried at 40°C for 40 minutes and re-suspended in 33 µl ddH₂O.

10 Tag Array Design and Hybridization

For each tag sequence, two probes were synthesized on the array. One is exactly the designed tag sequence (referred to as a Perfect Match, or PM probe). The other one is identical except for a single base difference in a central position (referred to as a Mismatch, or MM probe). The mismatch probe services as an internal control for
 15 hybridization specificity. Over 32,000 20-mer tag probes (and their companions) were chosen¹¹ and fabricated on a 8 mm x 8mm size of array. Each probe (feature) occupies a 30 microns x 30 microns area. The sets of arrays were synthesized together on a single glass wafer on which 100 arrays were made.

The labeled sample was denatured at 95°C - 100°C for 10 minutes and snap
 20 cooled on ice for 2 - 5 minutes. The tag array was pre-hybridized with 6 X SSPE-T (0.9 M NaCl, 60 mM NaH₂PO₄, 6 mM EDTA (pH 7.4), 0.005% Triton X-100) + 0.5 mg/ml of BSA for a few minutes, then hybridized with 120 µl hybridization solution (as shown below) at 42°C for 2 hours on a rotisserie, at ≈ 40 RPM. Hybridization Solution consists of 3M TMACl (Tetramethylammonium Chloride), 50 mM MES
 25 ((2-[N-Morpholino]ethanesulfonic acid) Sodium Salt) (pH 6.7), 0.01% of Triton X-100, 0.1 mg/ml of Herring Sperm DNA, 50 pM of fluorescein-labeled control oligo, 0.5 mg/ml of BSA (Sigma) and 29.4 µl labeled SBE products (see below) in a total of 120 µl reaction.

The chips were rinsed twice with 1X SSPE-T for about 10 seconds at room temperature, then washed with 1X SSPE-T for 15 - 20 minutes at 40°C on a rotisserie, at \approx 40 RPM. And then wash the chip 10 times with 6X SSPE-T at 22°C on a fluidic station (FS400, Affymetrix). The chips were stained at room temperature with 120 μ l staining solution (2.2 μ g/ml streptavidin R-phycoerythrin (Molecular Probes), and 0.5 mg/ml acetylated BSA, in 6 x SSPET) on a rotisserie for 15 minutes, at \approx 40 RPM. After staining, the probe array was washed 10 times again with 6 x SSPET on the FS400 at 22°C. The chips were scanned on a confocal scanner (Affymetrix) with a resolution of 60-70 pixels per feature, and two filters (530-nm and 560-nm, respectively). GeneChip Software (Affymetrix) is used to convert the image files into digitized files for further data analysis.

Clustering Analysis

For a given marker (at a given tag probe position), the intensity of each of the two colors (fluorescein and phycoerythrin) was calculated as the intensity at the perfect match position (PM) minus that at the mis-match position (MM). Negative fluorescein or phycoerythrin intensity values are treated as if they were zero. The Phat values were computed as the ratio of the intensities (fluorescein/fluorescein + phycoerythrin). The Phat values were sorted, and the optimal set of ranges for AA, AB and BB genotypes given the hypothesis of 2 or 3 clusters was considered, subject to the following rules: at most 4 points (outliers) may be excluded from the genotype ranges. For 2 groups, the total range Phat values must be at least 0.3. For 3 groups, the total range Phat values must be at least 0.5. Ranges must be separated by a gap of at least 0.1. The width of a range may be at most 0.4. A score was then computed as: $\text{Score} = 1 - (\text{sum of range widths} / \text{total range}) - (\text{outliers} * 0.1)$.

The set of ranges with the best score was found and used to call genotypes. This score increases with narrow ranges, while decreases with the number of points that are left out of any range. Therefore, it tends to be optimal when all the phat values are contained within relatively small ranges.

ABI Sequencing to Determine Genotypes

To independently confirm the genotypes called from the tag array assay, three samples (904957000000, 904896000000, and 904889000000) were sequenced using gel-electrophoresis based method. Samples were amplified for all sites with T7 and T3 tagged primers, using standard PCR cycling conditions (2.5 μ l of 20 ng/ μ l DNA, 0.375 μ l of 20 μ M primer (X2), 1.5 μ l of 10X PCR buffer, 0.9 μ l 25mM Mg^{2+} , 0.15 μ l 10mM dNTPs, 0.25 μ l 10 U/ μ l Taq DNA Polymerase (Sigma), brought up to 15 μ l with ddH₂O per tube). Some products were sequenced directly, while a M13 nesting strategy was used due to the close proximity of the polymorphic base to the primer end.

10 Samples from the initial amplification were diluted 1:50 with ddH₂O, and amplified with M13F-T7 (TGTAACGACGGCCAGTTAATACGACTCACTATAGGGAGA; SEQ ID NO: 9) and M13R-T3 (AACAGCTATGACCATGAATTAACCCTCACTAAAGGGAGA; SEQ ID NO: 10) primers using standard PCR conditions. All PCR products were cleaned with

15 Exonuclease I (Amersham 0.15 μ l of 10 U/ μ l per well) and Shrimp Alkaline Phosphatase (Amersham, 0.30 μ l of 1 U/ μ l per well) in a volume of 10 μ l. Dye terminator sequencing using a M13R primer (AACAGCTATGACCATG; SEQ ID NO: 11) or T7 primer (TAATACGACTCACTATAGGGAGA; SEQ ID NO: 12) on an ABI377 (Perkin Elmer) using Big Dyes (Perkin Elmer) was performed to determine the

20 genotype status for each SNP in all three individuals. Trace files were read with Edit View 1.0 (Perkin Elmer) software.

EXAMPLE 1

DNA from an individual is isolated, and amplified with primers from 15 previously-characterized (i.e., known) SNPs. Amplification is allowed to proceed as

25 described in Hudson, T.J. *et al.* (Science 270:1945-1954 (1995)) and Dietrich *et al.* (Dietrich, W. F. *et al.*, Nature 380:149-152 (1996); Dietrich, W. F. *et al.*, Nature Genetics 7:220-245; Dietrich, W. *et al.*, Genetics 131:423-447 (1992)). For example, in a 50 μ l reaction volume, 0.5 ng of template nucleic acid/target polynucleotide is added

to 1 μ M forward amplification primer, 1 μ M reverse amplification primer, 200 μ M dGTP, 200 μ M dTTP, 200 μ M dATP, 3.5 mM MgCl₂, 1.0 mM Tris-HCl (pH 8.3), 50 mM KCl, 0.02 μ M molecular probe, and 0.25 units of polymerase enzyme. The reaction mixture can then be subjected to a two-step amplification process, performed
5 on a Tetrad (MJ Research, Watertown, Massachusetts), with the conditions: denaturation at 94°C for 60 seconds, followed by an annealing/extension step at 53°-56°C for one minute. The denaturation and annealing/extension steps are repeated for 40 cycles. Alternatively, a three-step thermocycling reaction can be used, such as 94°C for 60 seconds, followed by annealing at 53°-56°C for 30 seconds, followed by
10 extension at 72°C for one minute the three steps being repeated for 40 cycles. This may be followed by an optional extension step at 72°C for five minutes.

After amplification is complete, locus-specific tagged oligonucleotides specific for the 10 SNPs are added, and are allowed to hybridize to the amplification products.

Reagents for a single base extension reaction are then added, where each of the
15 four ddNTPs is labeled with a different fluorophore. Single base extension is then performed as described by Kobayashi et al. (Mol. Cell. Probes 9:175-182 (1995)).

After the reaction is complete, the reaction products are placed in contact with the universal array, and the reaction products allowed to hybridize, each product to its appropriate oligonucleotide tag on the array. The chip is then assayed in a fluorometer,
20 and the wavelength emitted at each address in the array is recorded. From this data, the genotype at each individual SNP is determined.

EXAMPLE 2

Two alleles of template were mixed at ratios of 1:30, 1:10, 1:3, 1:1, 3:1, 10:1, and 30:1. These were labeled with different color labels by single-base extension
25 reaction and hybridized to a tag array. A correlation was observed between the signal intensity ratio and the template concentration ratio over a 900-fold dynamic range. See Figure 2.

EXAMPLE 3

A set of tag sequences is selected such that the tags are likely to have similar hybridization characteristics and minimal cross-hybridization to other tag sequences. An oligonucleotide array of all of the tags is fabricated. The design and use of such a
5 4,000-20mer-tag array for the functional analysis of the yeast genome has been described (1). More recently, Affymetrix designed and fabricated an array with a set of more than 16,000 such tags. The tag sequence synthesized on the chip can be 20-mer, 25-mer, or other lengths.

EXAMPLE 4

10 Marker specific primers are used to amplify each genetic marker (*e.g.* SNP). A multiplex PCR strategy is used to amplify these markers from genomic DNAs of tested individuals (2). After PCR amplification, excess primers and dNTPs are removed enzymatically. These enzymatically treated PCR products then serve as templates in the
15 next SBE reaction. Please note that these templates (PCR products) are double stranded, which are different from the templates used in other protocols (3, 4). For example, in Minisequencing (3) and Genetic Bit Analysis (GBA, 4), a double stranded template has to be converted to a single stranded template prior to the base extension reaction. The methods used for this conversion are costly, laborious, and hard to automate.

20 EXAMPLE 5

In the protocol described below, an SBE primer is designed for each genetic marker which terminates 1 base before the polymorphic site. However, other primer design schemes can be used. The primer for each marker is tailed with a unique tag which is complementary to a specific probe sequence synthesized on the tag chip. The
25 extension reaction is multiplex, in which SBE primers corresponding to multiple markers were added in a single reaction tube, and extended in the presence of pairs of

ddNTPs labeled with different fluorophores, *e.g.* for an A/C variant, there might be a ddATP-red and DDCTP-green.

EXAMPLE 6

5 The resulting mixture is hybridized to the tag array. Each tag corresponds to a single marker. The ratio of the intensities of the colors indicates the genotype (or the allele frequency, ranging from 0% to 100%) of the samples tested.

EXAMPLE 7

SBE template preparation: Marker specific primers are used to amplify each single nucleotide polymorphism (SNP). A multiplex PCR strategy is used to amplify
10 these SNPs (Science 280:1077-1082, 1998).

Multiplex PCR:

Multiplex PCR reaction is carried out with AmpliTaq Gold and 25 primer pairs in a 25µl reaction volume. SNPs with same base composition at the polymorphic site (i.e. A/G, T/C, etc) are pooled together.

15 **PCR reagents:**

10XPCR Multiplex Buffer (II): 100 mM Tris/HCl (pH 8.3)
500 mM KCl

25 mM dNTPs

20 F & R Primers (for each primer, the conc. is 1 µM)
20 ng/µl Genomic DNA

Multiplex PCR reaction (25 µl)

Primer Mix (1 µM each) 2.5 µl
Genomic DNA (20 ng/µl) 2.5 µl

	10XPCR Buffer II	2.5 µl
	25 mM MgCl ₂	5 µl
	25 mM dNTPs	1 µl
	AmpliTaq Gold (5U/µl)	0.4 µl
5	ddH ₂ O	up to 25 µl

PCR conditions

96°C 10 min

40 cycles :

	94°C	30 sec
10	57°C	40 sec
	72°C	1 min 30 sec
	72°C	10 min
	4°C	O/N

Enzymatic treatment of PCR products to degrade and de-phosphorylate the unused
15 primers and dNTPs, respectively:

To a 25 µl PCR products, add 1 µl of Exonuclease I (Amersham Life Science,
10 U/µl) and 1 µl of Shrimp Alkaline Phosphatase (Amersham Life Science, 1 U/µl),
and incubate at 37° C for 1 hour. Inactivate the enzyme activities at 100°C for 15
minutes. Apply the sample to a S-300 column (Pharmacia), to further reduce the
20 residual PCR primers and dNTPs, and replace the buffer with ddH₂O. The sample is
ready for next SBE reaction.

Single Base Extension (SBE):

An SBE primer is designed for each SNP which terminates 1 base before the polymorphic site. The primer for each SNP is tailed with a unique tag which is complementary to a specific probe sequence on the tag chip. The SBE reaction is also multiplexed at 25-plex.

Reaction Mixture (33 μ l):

	Template (see above)	6 μ l
	SBE Primer mix (20 nM for each primer)	2.5 μ l
	5X Thermo Sequenase buffer	6.6 μ l
10	Bio-(d)dNTP(X nmol/ μ l*, NEN)	0.5 μ l
	Flu-ddNTP(1nmol/ μ l, NEN)	0.8 μ l
	Other two cold –ddNTPs(1nmol/ μ l, Biopharmacia)	0.3 μ l each
	Thermo Sequenase(6.4 U/ μ l) (Amersham)	0.4 μ l
15	ddH ₂ O	up to 33 μ l

* X= 0.5 when it is Bio-ddUTP or bio-dCTP(0.5 mM), or X= 0.25 when it is bio-ddATP (0.25 mM)

PCR program:

	96°C	3'	1 cycle
20	94°C	25"	
	58°C	11"	45 cycles
	4°C	forever	

Precipitation:

After SBE reaction, we combined 9 tubes for each sample, mix with 30 µl of 100 µg/ml glycogen (Boehringer Mannheim), then precipitated with 18.75 µl of 8 M LiCl, and 1125 µl of pre-chilled (-20°C) ethanol (Abs.). Mix well; then centrifuge at the top speed (Eppendorf centrifuge 5415C) for 15 min at room temperature; Decant the supernatant, and dry the samples at 40C for 40 min, re-suspend the samples in 33 µl ddH2O, now it is ready for hybridization.

Hybridization:

The prepared sample is denatured at 100°C for 10 minutes and snap cooled on ice for 2-5 minutes. The universal tag chip is pre-hybridized with 6 X SSPE-T (0.9 M NaCl, 60 mM NaH₂PO₄, 6 mM EDTA (pH 7.4), 0.005% Triton X-100) + 0.5mg/ml of BSA, then hybridized with 120 µl hybridization solution (as shown below) at 42°C 2 hours on a rotisserie, at ≈ 40 RPM.

The hybridization solution contains:

15	5M TMACL	72 µl
	0.5M MES (pH 6.7)	12 µl
	1% Triton X-100	1.2 µl
	HS DNA (10mg/ml)	1.2 µl
	Flu-c213 (5 nM)	1.2 µl
20	BSA (20 mg/ml)	3.0 µl
	Plus 29.4 µl prepared sample (see above).	

Post-Hybridization Wash:

Rinse the chip with 1X SSPE-T 10" twice first, then wash with 1X SSPE-T for 15-20min at 40°C on a rotisserie, at ≈ 40 RPM. And then wash on a fluidic station (FS400, Affymetrix) 10 times with 6 x SSPET at 22°C.

Staining:

Stain the chip at room temperature with 120 μ l staining solution (2.2 μ g/ml streptavidin R-phycoerythrin (Molecular Probes), and 0.5 mg/ml acetylated BSA, in 6 x SSPET) on a rotisserie for 15 minutes, at \approx 40 RPM. After staining, the probe array was
 5 washed 10 times again with 6 x SSPE-T on the FS400 at 22°C.

Scanning:

The chips were scanned on a confocal scanner (Affymetrix) with a resolution of 60-70 pixels per feature, and two filters (530-nm and 560-nm, respectively). GeneChip Software (Affymetrix) is used to convert the image files into digitized files for further
 10 data analysis.

EXAMPLE 7

Genotyping With High-Density Oligonucleotide “Tag” Arrays

A genotyping method based on the use of a high-density “tag” array that contains over 32,000 pre-selected 20-mer oligonucleotide probes, combined with
 15 marker-specific PCR amplifications and single base extension (SBE)¹⁻² reactions has been developed. We have used this method to genotype a collection of 144 single-nucleotide polymorphism (SNPs) identified from 49 hypertension candidate genes³. First, marker-specific primers were used in multiplex PCR reactions to amplify specific genomic regions containing the SNPs. The PCR amplified DNA products were
 20 then used as templates in SBE reactions. Each SBE primer comprises a 3' portion and a 5' portion. The 3' portion is complementary to the specific SNP locus and terminates one base before the polymorphic site. The 5' portion comprises a unique sequence, which is complementary to a specific oligonucleotide probe synthesized on the “tag” array. The extension reaction is multiplex, with SBE primers corresponding to multiple
 25 SNPs in a single reaction tube. The primers are extended in the presence of two-color labeled ddNTPs, and the resulting mixture is hybridized to the tag array. The intensity ratio of the two colors was used to deduce the genotypes of the samples tested.

The tag array strategy begins with an array of tag sequences selected in a manner that all tag probes are in the same length, *e.g.* 20-nucleotide long, with similar melting temperature and G-C content, and the lowest sequence homologous among each other¹¹. Therefore, these tags are likely to have similar hybridization characteristics and minimal cross-hybridization to other tag sequences.

The design and use of a 4,000-tag array for the functional analysis of yeast *Saccharomyces cerevisiae* genes¹¹ and drug sensitivity studies¹² have been described. More recently, we have designed and fabricated an array that contains more than 32,000 such tags, and developed it as a genotyping tool, in combination with marker-specific PCR amplifications and SBE reactions.

As shown in Fig. 7, marker specific primers are designed and used to amplify each single nucleotide polymorphism (SNP). A multiplex PCR strategy is used to amplify these SNPs from genomic DNAs⁹. In general, SNPs with same base composition at the polymorphic site (*e.g.* all the A/G polymorphisms) are grouped together. After PCR amplification, excess primers and dNTPs are degraded and de-phosphorylated using Exonuclease I and Shrimp Alkaline Phosphatase, respectively. These enzymatically treated PCR products (double-stranded) are then served as templates in the SBE reaction. A SBE primer is designed for each genetic marker, which terminates one base before the polymorphic site. Each primer is tailed with a unique tag that is complementary to a specific probe sequence synthesized on the tag array. The extension reaction is multiplex, in which SBE primers corresponding to multiple markers (up to 56 markers that we have tested so far) were added in a single reaction tube, and extended in the presence of pairs of ddNTPs labeled with different fluorophores, *e.g.* for an A/G variant, biotin-labeled ddATP and fluorescein-labeled ddGTP are used. The resulting mixture of SBE reactions is hybridized to the tag array. Each tag hybridizes to a specific probe position on the chip. The ratio of the intensities of the colors indicates the genotype (homozygous wild type, or homozygous mutant, or heterozygous) or the allele frequency (ranging from 0% to 100%) in the samples tested.

In a comparison of the results of using single-stranded and double-stranded PCR products as the templates in the current SBE/tag array assay, no significant difference was found (data not shown). However, in previously published protocols of minisequencing¹³⁻¹⁵ and genetic bit analysis¹⁶⁻¹⁸, a double-stranded template has to be converted to a single-stranded template prior to the base extension reaction. The methods used for this conversion were costly, laborious, and hard to automate.

The tag array assay provides a fairly accurate quantitative measurement of the allele frequency in samples tested. As shown in Figure 2, we have synthesized two artificial SBE templates. They are identical, except the 21st position: T in template-T, and G in template-G. We then mixed the two templates at ratios of 1:10, 1:3, 1:1, 3:1, 10:1, and 30:1, which is a 300-fold dynamic range. Six SBE primers, which have the same 3' portion (the portion complementing to the template sequence) but different 5' portion (the portion complementing to the tag probes on the tag arrays) were designed (Figure 2), and extended in the presence of the SBE templates mixed at different ratios, and biotin-labeled ddATP and fluorescein-labeled ddCTP. As shown in Fig. 8, the intensity ratio of the two colors and the template concentration ratio (i.e. the allele frequency) appears to form a fairly good linear correlation in the 300-fold dynamic range that we tested.

To further test the robustness and the efficiency of the tag array/SBE assay method for genotyping application, we set out to type a portion of the SNPs that we had identified from a large-scale polymorphism screening study with the hypertension candidate genes³. Initially, we selected 173 SNPs from 56 hypertension candidate genes. These SNPs were chosen for their being occurred in promoter regions, or splicing junctions, or coding regions in which the nucleotide changes caused amino acid changes. We reason that these SNPs can be the good candidates for being the functional mutations predisposed to the disease. Therefore, the assay developed in this study could then be used in large-scale association studies in hypertension. PCR primers were designed and tested individually for these 173 SNPs. 8 of them (4.6%) failed to amplify. SBE primers were then designed for the remaining 165 SNPs. A multiplexing PCR and

5 multiplexing SBE assay was developed with a complexity of 9 to 28 markers in each reaction and a total of 9 reactions for the 165 markers. 21 of them (12.7%) failed in the multiplexing PCR and multiplexing SBE assay. Therefore, 144 markers from 49 genes passed the assay development. The gene location, polymorphic sites, and the designed primers for these 144 markers were summarized in Table 1.

We then genotyped 44 individuals using 44 tag arrays. Good hybridization signals were obtained in 96.5% (6116 / 6336 (144 x 44)) of the cases. The signal intensity values from the hybridization results were used in clustering analysis for each of the 144 markers. Genotypes for each individual at the 144 loci were assigned automatically based on the clustering results, with some manual editing. Data Desk 6.0 (Data description, Inc.) was used to manually display the clustering analysis results (of the intensity ratios of the two colors). Overall, 80-85% of the markers form good cluster(s).

15 We have performed the gel-based DNA sequencing to determine the genotypes at 115 loci in 3 of the 44 individuals (see Methods). Comparison of the ABI sequencing results and the chip results resulted in 14 discrepancies (4%), out of $115 \times 3 = 345$ genotype calls. Most of the discrepancies occurred in cases where one method called homozygous, while the other method called heterozygous. In one case (marker ICAM1ex6.254), where the ABI sequencing method called G/G, but the tag array /SBE assay method called A/A in all the three individuals, we believe the discrepancies are due to mis-priming of the SBE primer to adjacent sequences.

25 We also tested the reproducibility of the tag array/SBE assay genotyping method. We repeated the multiplexing PCR, SBE and the chip hybridization experiments in 4 individuals. The ratios of the two colors (for each of the 144 markers) in the replicated experiments are not all exactly the same, but they all fall into the same cluster (i.e. giving the same genotype call). Therefore, we didn't find any discrepancy in the genotyping call of duplicated samples.

Table 1

[illegible]

00460074394660

Table 1 (cont)

AT2EX3 807	GGGAGAGACAG/AATACCCGT	GACGAATAGCTTGGGAAGAC	ACTGTGACGGGTATCC	TGGCATCTGTTTCCCTGTGGACGAATAGCTATGGGAAGACA
AVPEX2 154	AATACCTGCTGCTGCGCCCTGC	TGGCAGGAGAGACATACCT	GGATGGCAGGGCA	TATCATCTGTGCTGCGCCGACGAGGAGAACTACCTGC
AVPR2EX2 129	CGGGAGCTGGCTGCTGCTGCTGC	CTAGCCGGGGGGGAG	CCACAAGACATATGGAGAGACAG	GTGGATCTTATGTAATAGCTAGCCGGGGGAGAGCTGG
AVPR2EX2 444	CCATGCTGGCTGCTGCTGCTGCTGC	TGCGCTCCCATGCTG	CTACTCCATATGGCGTA	GGCTCAATGGGGTCTGTAATATGCTGCGCTCCCATGCTGG
BIK 1521	GGCAGCTTTGAGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	CTACTCCATATGGCGTA	GGCAGCTATGCTGCTGTAATAGAGTGGTGGGACCTTTGA
BIK 2463	TACCTGGGCTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTATGCTGCTGTAATAGAGTGGTGGGACCTTTGA
BR3EX1 730	CATCATATATACCTTATGCTGT	GAAAGCATATGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	ATCTAGAGTGGATCGGATAGAGTGGTGGGACCTTTGA
CAL/CRPEX4 30	TCCCTGGAGCTGCTGCTGCTGCTGC	CTAGCTTTGAGCTGCTGCTGCTGC	CTAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
CaR AA1011	GACCGACACCTGAGCCATACCT	CTAGCTTTGAGCTGCTGCTGCTGC	CTAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
CaR AA990	CATGGCCCTGAGTGGGATCTAC	CTAGCTTTGAGCTGCTGCTGCTGC	CTAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
CLNKBEX10 33	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
CLNKBEX15 64	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
CLNKBEX4 19	GGATATCAAGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
COX2EX1 358	CCAAATGTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
COX2EX10 156	ATGGTAGAGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
CYP11BEX4 205	AGGAGGACTTGTGGAGGCTGGG	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
CYP11BEX5 107	GTGGGCGGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
CYP11BEX3 152	CAGCCCTGAGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
CYP11BEX6 91	GACGAGTGGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
CYP11BEX7 65	GTGAGCTGAGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
DBH1EX4 132	GATGAACCCCTGAGTGGGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
DBH1EX5 39	AGCCGCTTGTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
DDIR122	CTAGAGGAGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
EDNRBEX3 144	GATATATATCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
ELAM1 77	GACTTGTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
ELAM1EX5 197	TGGGAGACAGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
ELAM1EX7 200	GTGGGAGACAGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
ENOS 78	CCCAGATGATGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
ET1EX5 90	TGAAGGAGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
GALNREX1 327	GCAGGAGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
GALNREX1 553	TCAGAGGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
GGREX9 29	AACATGGCTTGTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
GLUT2EX1 137	AGCACTAATGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
GLUT2EX1 164	CAGTGTGCTTGTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
GLUT4EX3 112	GGCAGCTTGTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
GMP-140 105	TTTCTGCTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
GMP-140 164	CCAGTGGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
GMP-140 25	TACTCCAGGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
GMP-140 30	GAGCCGCTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
GNB3EX10 144	CATCATCTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
GSV1EX16 210	CATCCGCTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
GSV1EX3 117	GGAGGCTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
GSV1EX8 43	AACGAGGCTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
HUMGUANCYC 2905	ATGTGGGCTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
ICAM1 117	TTCCTGGAGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
ICAM1EX1 683	TTCGAGCTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
ICAM1EX2 115	GACGAGCTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
ICAM1EX6 254	GATGAGCTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
ICAM2EX3 63	AAGAGAGCTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC
IRS-2 AA1057	CAGGAGCTGCTGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GGCAGCTTTGAGCTGCTGCTGCTGC	GTGCTGCTGCTGCTGCTGCTGCTGCTGC

bioRxiv preprint doi: <https://doi.org/10.1101/2016.05.04.050000>; this version posted May 4, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1 (cont)

KLKX3 253	GTTCGCCACC(C/G)AGGAACCCGA	TGCTGGAGTGGCCAC	CCCCATTGGGTTCC	CTGTGCGCCCTGATAGAAATGTGTGGAGTTGGCCACC
KLKX4 110	GTCCAGAGGGT(A/G)ACAGACTTC	AAGCCACAGTCCAGAAAG	CGACACAGCATGAAGTCTG	ACGCCAATATGGCCATCTGTGGCAAAAAGCCACGTCGAGAAGG
LAMI 103	CCAGTGTACAGT(C/T)TTGTGAAGTC	TGGGCCCCAGTGTC	GCAGAAAGAAAGAGAGACTTACC	CTGTGCCCTGCTGTGATGATTACTATGTGGCCCCAGGTGTGAG
LPL 177	TATGAGATCA(A/G)TAAGTCAGA	CAACAATCTGGGCTATGAGATC	TGCTCTTTTGGCTCTGACTT	GTGGCTGTGACATATAGTACATCTGGGCTATGAGATCA
LPL 98	AATAAGAAAGT(C/G)AGGCTGGTGA	CCATGAAGATGCTCTGAATAAGAA	CCAGAATGCTACCAAGCC	CCGTATGTGAGTCTCTGACGCATGACAAGATCTCTGAAATAGAAAT
MACEX13 R NA	TGCTGGTCC(C/T)AGCAGGAGG	GCACCTCTGCTGTGCC	TGACTGTACCTGTGTGGA	CACATCTGACACCGTATGTTGGATGCTCTCTGGCT
MRLEX2 345	CATGGCCGCC(A/G)TGTGTAAGG	GGTGGCTGATCGGC	CACATGATAGGCTTTTAACT	GGATGTCTGCCGTGGCAATATGTGTGGCTATGGCGCC
NAT2 346	CAGGTGACCA(A/G)TGAGCGGACGG	CCCTCTCTGGAGGTGACC	ACAATGTAATTCCTGGCGTC	CTGTGAGTATGTACGGTCTCTCTCTGCAAGGTGACCA
NETEX11 123	AGCTCTGCC(TT)GCTCTCTGGTG	GAAGTGTGAGTCCGTC	CGCTGGAGCATACACACC	GGCTGGGTTCTCATCTGATCTGGAAATCTGTCAGTCTGGCT
NETEX12 81	CTAGACGAGT(G/A)CATCTTCCC	GCACCTCACTACGAGGA	CCAGGCGGGAAGATG	CGGCTGGGTAGCACTAATAGCCACCTACCTGAGGAC
NETEX13 121	AATGGCAATG(A/G)TACTCTCTG	GAGCTCCATAGGCATC	GTCCATGTGGAGTGGC	GCATGAAGTTCCTAATCTGGGAGCTCTCAATGGCATCA
NETEX7 112	TGGTACATG(C/C)CCATGAACA	TCCTCTCTCTTGGTTACA	TGTTGACTGTGTTCATGG	CGGACATATGCTGATAGTCTGATCTCTCTCTGTTTACATG
NETEX7 131	CACAAAGTCA(A/G)ATGTAGGAT	GCCCATGACACAAAGTTC	TGTGGCCATCTCTCAAT	CGGCAATATGATATAGTCTGATGCTCCCATGACACAAGTCA
NETEX7 73	CACCAAGTTC(G/G)TCTCTGGGT	GCCTAACCTGTATACCAAGCTT	AGATGGCAACCAAGAG	CCCTGTATGATGAGTGGAGTCTAGGATCACTGATATACCAAGTTC
NETEX9 157	TGCTAAGCA(A/G)GTGTGATAGG	CGCCCTGTCTGCTAAGC	GCACAGCTCTGCTAC	CCAAAGATGCTACTGTAGTCTGAGTCTGCTCTGCTGCTGCTG
OB 160	GATCAATGAC(A/G)TTTACACAC	AATGTGACAGGATCAATGA	ACCTCTTACCTGCTGTGAA	CATTTGACCCACTGAGATGGATTTGTACAGGATCAATGAC
OB-R 174	GTAAATTTTCC(A/G)TTCACCTCTA	TCACATCTGTGGAGTAAATTTTC	GTGAACTGACATAGAGGTGA	CACGGATCTGCCGCTGGAATGATCTGTTGGAGTAAATTTTCC
PGISEX1 396	GGGAGCAGGG(T/G)TTCTCCAGA	GCTGGGGAGGAGG	GGGCTCTGTGGAGA	CGAAACATATGCGGTGGAATAAGTCTGGGAGGACGGG
PLA2AEX2 42	GCCTGCCCG(A/G)CAGCGGCATC	CTTGCACTGGCCGCC	AGGGCTGATCGGCT	AGATGAGTCTGATGCGGCTTTGCAAGTGGCCGCCCG
PLA2AEX3 104	TGCTGGACAA(G/A)CCGTACACCC	TGGACAGCTGTAAATTTCTGCT	ATGAATAGGTGTGGGTGTACG	TGCTCTATTTGTGACTATGGACAGCTGTGTAATTTCTGTGGACAA
PNMTX3 181	GGAGGCTGTG(A/T)GCCAGATCT	CCCTTCTGCTTGGAGGCTGT	AGCTGGCAAGATCTGGG	TGTGAGCTTTGTACTACGGCTGCTTCTGCTTGGAGGCTGTG
PNMTX3 251	GGGGGGCACC(T/A)XCTCTCTATC	GCTGAGGCTTGGGG	CCAGGTACCAAGACTCTCTC	TGTGAATAATGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
PNMTX3 269	ATCGGGGCCCTT(A/G)AGGAGGATCG	GCTGAGGCTTGGGG	CCAGGTACCAAGACTCTCTC	TGAGCAATTTAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
PON2 384	CCCTACTTAC(A/G)ATCTGGAG	TCACTATTTCTTGGCCCTACTT	CCAAATACATCTCCCAAGGA	GATCGCAATTTAGAGGCTGATATTTCTGACCCCTACTATAC
SCNNIB 222	AACATCTAT(G/C)TGAGAGGCT	CCGCATCCAGAACTTCTA	CATAACTGTAGTCACTGTAGGCTTCT	CAGTCTGTGGATAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
SCNNIB 238	CACCTGAC(A/G)TGCCGCCGCC	GGAGGCCACACCAACTT	GTGTGGGCTGCCG	GACTGGGATTTACATGCTATGAGGCTGCTGCTGCTGCTGCTGCTGCTG
SCNNIB AA442	ACCTGCAATG(G/T)CATGTGCAAG	CGCAGAGAGAGACTGCAAT	GCAGAGCTCTTGACAT	GCACCGTCTGTGCACTATACAGAGAGAGAGAGAGAGAGAGAGAG
SCNNIG 172	GGCTCCCAAT(G/G)TCCAGAAAG	GGAAACAGGCTCTCCCC	ACGGGAGCTTCTGGA	AGCCAAAGTGCAGGCTGATCATCTGGAACACAGGCTCTCTGCTGCTG
SCNNIG 21	TTCTGTCCAA(T/G)TCCGTTGGCC	CAGTATGAGATGCTTCTGTCCA	CCAGCTGGCCACCGA	TGCTCTGCTTGGATGTGAGCAGTATTTGAGATGCTTCTGCTGCTGCTG
SCNNIGEX1 216	GTGTGGCC(G/T)TCCGGGCGG	CGTTGTGAAGTCTGTGCGC	CTGAGCCGCCCGGA	CAGTGAAGTGTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
SCNNIGEX2 219	GGTGTCTCCG(G/T)TCCGCTGCGG	GCATGTGGTGTGCGG	GGAGGCGGCGCAGAC	CTCAGCAGTATGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
SCNNIGEX3 259	GGGAAAGT(G/G)AGGCTGTAGAT	GGGAGGAGGCGGAAAGT	GAAGCTTGTGAATGATGCT	CTTATGGCGCTGTGGCTATCAGGGAGGAAAGCGGAAAGTCTG
TBXASEX11 88	CCCCGAGGCT(G/A)CTGTGTAGA	CGAGGTCTGGGACA	ACGGCATCTGACACA	GATATGCTTACGTGAGTCTGCGGCACTCTGACACAG
TBXASEX9 276	TGCCACCTAG(G/G)TACTGGCCAC	CACACTTCTTTTGGCCACT	AGGGTGTGGGCGAGT	CAACAACCTGGCGAGATGAAACACACTTCTTTTGGCCACTAC
TGF-B1 75	TCATGGCCCA(T/G)CCGCTGGAG	TCCTGCTCTCATGGCC	GGGCTCTCCAGCGG	TTGTGCAATTTGGAGCGGCTTCTGCTGCTTCTGCTGCTGCTGCTG
TRIREX1 56	GCAGAACTTA(G/C)ATGATAAGCA	CAGGTACTAGAGTCTTCCAGAACTT	GGCTTTGCTTGTCTTACA	AGCAGTAAATGACAGGCTGCAAGGTACTAGAGTCTTCTGAGAACTTA

References:

1. Singer-Sam & Riggs, *Methods Enzymol.* 225:344-351 (1993).
2. Greenwood & Burke, *Genome Res.* 6:336-348 (1996).
3. Halushka *et al.*, Pattern of single nucleotide polymorphisms in human genes.
- 5 *Nature Genet.* (Submitted).
4. Risch & Merikangas, *Science* 273:1516-1517 (1996).
5. Collins *et al.*, *Science* 278:1580-1581 (1997).
6. Collins *et al.*, *Science* 282:682-689 (1998).
7. Chakravarti, *Nature Genet.* 21:56-60 (1999).
- 10 8. Chee *et al.*, *Science* 274:610-614 (1996).
9. Wang *et al.*, *Science* 280:1077-1082 (1998); <http://www.genome.wi.mit.edu/SNP/human/index.html>.
10. Lipshutz *et al.*, *Nature Genet.* 21:20-24 (1999).
11. Shoemaker *et al.*, *Nature Genet.* 14:450-456 (1996).
- 15 12. Giaever *et al.*, *Nature Genet.* 21:278-283 (1999).
13. Pastinen *et al.*, *Clin. Chem.* 42:1391-1397 (1996).
14. Pastinen *et al.*, *Genome Res.* 7:606-614 (1997).
15. Pastinen *et al.*, *Hum. Mol. Genet.* 7:1453-1462 (1998).
16. Nikiforov *et al.*, *PCR Methods and Applications* 3:285-291 (1994).
- 20 17. Nikiforov *et al.*, *Nucleic Acids Res.* 22:4167-4175 (1994).
18. Head *et al.*, *Nucleic Acids Res.* 25:5065-5071 (1997).
19. Tobe *et al.*, *Nucleic Acids Res.* 24:3728-3732 (1996).
20. Delahunty *et al.*, *Am. J. Hum. Genet.* 58:1239-1246 (1996).
21. Chen *et al.*, *Genome Res.* 8:549-556 (1998).
- 25 22. Lyamichev *et al.*, *Nature Biotech.* 17:292-296 (1999).
23. Newton *et al.*, *Nucleic Acids Res.* 17:2503-2516 (1989).
24. Lo *et al.*, *Nucleic Acids Res.* 19:3561-3567 (1991).
25. Zhang *et al.*, *Proc. Natl. Acad. Sci. USA* 89:5847-5851 (1992).

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2

CLAIMS

What is claimed is:

1. An oligonucleotide array comprising one or more oligonucleotide tags fixed to a solid substrate, wherein each oligonucleotide tag comprises a unique known arbitrary nucleotide sequence of sufficient length to hybridize to a locus-specific tagged oligonucleotide, wherein the locus-specific tagged oligonucleotide has at its first end nucleotide sequence which hybridizes to, e.g., is complementary to, the arbitrary sequence of the oligonucleotide tag.
2. A kit comprising:
 - (a) an array comprising one or more oligonucleotide tags fixed to a solid substrate, wherein each oligonucleotide tag comprises a unique known arbitrary nucleotide sequence of sufficient length to hybridize to a locus-specific tagged oligonucleotide; and
 - (b) one or more locus-specific tagged oligonucleotides, wherein each locus-specific tagged oligonucleotide has at its first (5') end nucleotide sequence which hybridizes to, e.g., is complementary to, the arbitrary sequence of a corresponding oligonucleotide tag on the array, and has at its second (3') end nucleotide sequence complementary to target polynucleotide sequence in a sample.
3. A method of genotyping a nucleic acid sample at one or more loci, comprising the steps of:
 - (a) obtaining a nucleic acid sample to be tested;
 - (b) combining the nucleic acid sample with one or more locus-specific tagged oligonucleotides under conditions suitable for hybridization of the nucleic acid sample to one or more locus-specific tagged

oligonucleotides, wherein each locus-specific tagged oligonucleotide comprises a nucleotide sequence capable of hybridizing to a complementary sequence in an oligonucleotide tag and a nucleotide sequence complementary to the nucleotide sequence 5' of a nucleotide to be queried in the sample, thereby creating an amplification product-locus-specific tagged oligonucleotide complex;

- (c) subjecting the complex to a single base extension reaction, wherein the reaction results in the addition of a labeled ddNTP to the locus-specific tagged oligonucleotide, and wherein each type of ddNTP has a label that can be distinguished from the label of the other three types of ddNTPs;
- (d) contacting the complex with an oligonucleotide array comprising one or more oligonucleotide tags fixed to a solid substrate under suitable hybridization conditions, wherein each oligonucleotide tag comprises a unique arbitrary sequence complementary and of sufficient length to hybridize to a complementary sequence in a locus-specific tagged oligonucleotide, whereby the complex hybridizes to a specific oligonucleotide tag on the array; and assaying the array to determine the labeled ddNTPs present in the complex hybridized to one or more oligonucleotide tags,

thereby determining the genotype of the queried nucleotide in the sample.

4. A method to aid in determining a ratio of alleles at a polymorphic locus in a sample, comprising the steps of:
- (a) using a pair of primers to amplify a region of a nucleic acid in a sample, wherein the region comprises a polymorphic locus, whereby an amplified DNA product is formed;
- (b) labeling an extension primer by a single base extension reaction to form a labeled extension primer, wherein the amplified DNA product is used as a template, wherein the extension primer comprises a 3' portion and a

5' portion, wherein the 3' portion is complementary to the amplified DNA product and terminates one nucleotide 5' to the polymorphic locus, wherein the 5' portion is not complementary to the amplified DNA product, whereby a labeled dideoxynucleotide which is complementary to the polymorphic locus is coupled to the 3' end of the extension primer, wherein each type of dideoxynucleotide present in the reaction bears a distinct label; and

(c) hybridizing the 5' portion of the extension primer to one or more probes complementary to the 5' portion which are immobilized to known locations on a solid support.

5. The method of claim 4 wherein two complementary strands of the amplified DNA product are present in the single base extension reaction.
6. The method of claim 4 wherein two complementary strands of the amplified DNA product are used as templates in the step of labeling.
7. The method of claim 4 wherein the label is a fluorescent label.
8. The method of claim 4 wherein the label is a radiolabel.
9. The method of claim 4 wherein the label is an enzyme label.
10. The method of claim 4 wherein the label is an antigenic label.
11. The method of claim 4 wherein the label is an affinity binding partner.
12. The method of claim 4 further comprising the step of:
 - (d) optically detecting a fluorescent label on the solid support.

13. The method of claim 4 wherein the step of labeling employs at least two distinct dideoxynucleotides bearing distinct labels.
14. The method of claim 4 wherein the step of labeling employs four distinct dideoxynucleotides bearing distinct labels.
- 5 15. The method of claim 4 further comprising the steps of:
 - (d) comparing quantities of a first and a second label at a location on the solid support; and
 - (e) determining the ratio of nucleotides present at the polymorphic locus in the sample.
- 10 16. The method of claim 15 wherein the ratio of nucleotides present at two or more polymorphic loci is determined simultaneously.
17. The method of claim 4 wherein the sample comprises DNA from two or more individuals.
18. The method of claim 17 wherein the ratio of nucleotides present at two or more
15 polymorphic loci is determined simultaneously.
19. The method of claim 4 wherein the solid support is selected from the group consisting of beads, microtiter plates, and oligonucleotide arrays.
20. A set of primers for use in determining a ratio of nucleotides present at a polymorphic locus, comprising:
 - 20 (a) a pair of primers which when in the presence of a DNA polymerase amplify a region of double stranded DNA, wherein the region comprises a polymorphic locus; and

- 5 (b) an extension primer which comprises a 3' portion which is complementary to a portion of the region of double stranded DNA and a 5' portion which is not complementary to the region of double stranded DNA, wherein the extension primer terminates one nucleotide 5' to the polymorphic locus.
21. A kit comprising in a single container two or more of the sets of primers of claim 20.
22. A kit comprising in a single container:
- 10 (a) a set of primers of claim 20; and
- (b) a solid support comprising a probe which is attached to a solid support, wherein the probe is complementary to the 5' portion of the extension primer.
23. The kit of claim 22 wherein the solid support is an oligonucleotide array.
24. The kit of claim 22 wherein the solid support is a bead.
- 15 25. The kit of claim 22 wherein the solid support is a microtiter plate.
26. A method to aid in determining a ratio of alleles at a polymorphic locus in a sample, comprising the steps of:
- 20 (a) labeling an extension primer by a single base extension reaction to form a labeled extension primer, using a DNA molecule as a template, wherein the extension primer comprises a 3' portion and a 5' portion, wherein the 3' portion is complementary to the DNA molecule and terminates one nucleotide 5' to a polymorphic locus, wherein the 5' portion is not complementary to the DNA molecule, whereby a labeled

dideoxynucleotide which is complementary to the polymorphic locus is coupled to the 3' end of the extension primer, wherein each type of dideoxynucleotide present in the reaction bears a distinct label; and

- 5 (b) hybridizing the 5' portion of the extension primer to one or more probes complementary to the 5' portion which are immobilized to known locations on a solid support.

27. The method of claim 26 wherein two complementary strands of the DNA molecule are present in the single base extension reaction.

10 28. The method of claim 27 wherein each complementary strand of the DNA molecule is used as a template to label an extension primer.

29. The method of claim 26 wherein the label is a fluorescent label.

30. The method of claim 26 wherein the label is a radiolabel.

31. The method of claim 26 wherein the label is an enzyme label.

32. The method of claim 26 wherein the label is an antigenic label.

15 33. The method of claim 26 wherein the label is an affinity binding partner.

34. The method of claim 26 further comprising the step of:

- (c) optically detecting a fluorescent label on the solid support.

35. The method of claim 26 further comprising the steps of:

- 20 (c) comparing quantities of a first and a second label at a location on the solid support; and

(d) determining the ratio of nucleotides present at the polymorphic locus in the sample.

36. The method of claim 35 wherein the ratio of nucleotides present at two or more polymorphic loci is determined simultaneously.

5 37. The method of claim 26 wherein the sample comprises DNA from two or more individuals.

38. The method of claim 34 wherein the ratio of nucleotides present at two or more polymorphic loci is determined simultaneously.

10 39. The method of claim 26 wherein the step of labeling employs at least two distinct dideoxynucleotides bearing distinct labels.

40. The method of claim 26 wherein the step of labeling employs four distinct dideoxynucleotides bearing distinct labels.

An array of oligonucleotides on a solid substrate is disclosed, which can be used for multiple purposes. Methods and reagents are provided for performing genotyping to determine the identity or ratio of allelic forms of a gene in a sample. A single base extension primer is coupled to a sequence identity code. During the primer extension reaction a distinctive label is incorporated which identifies the allelic form present in the sample. This permits multiple simultaneous analyses to be performed easily and efficiently.

GENERIC HYBRIDIZATION ARRAYS
SOLID SUBSTRATE WITH GENERIC
MOLECULAR ADDRESSES

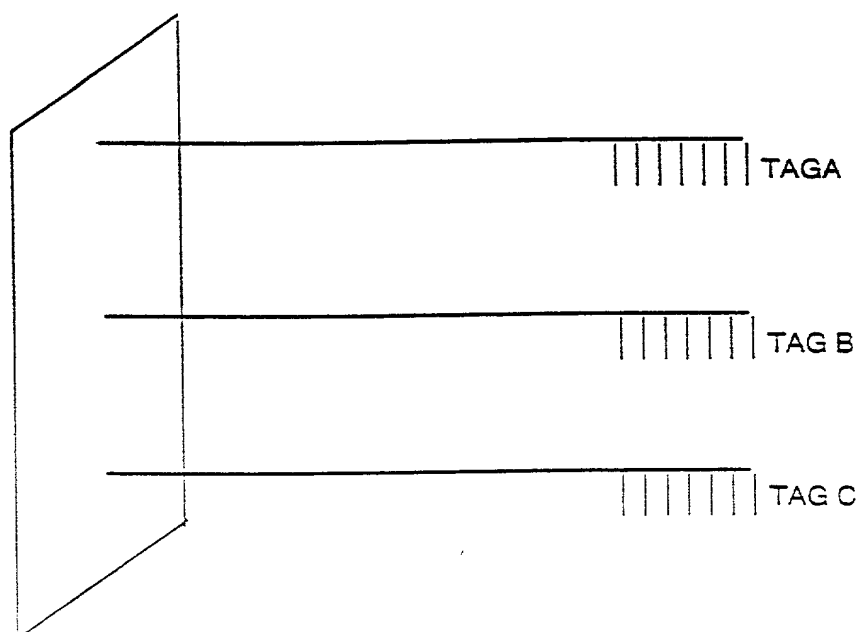


Fig. 1

Locus-specific tagged oligonucleotide:

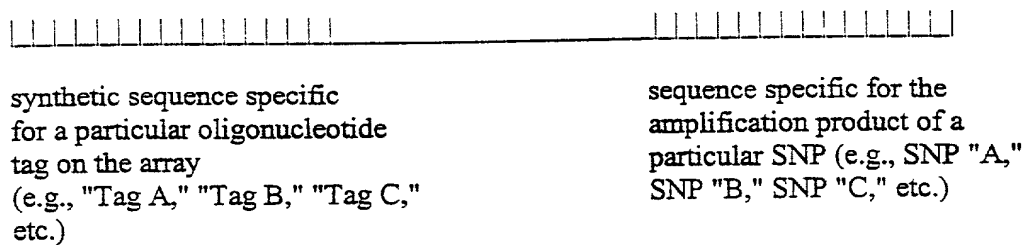


Fig. 2

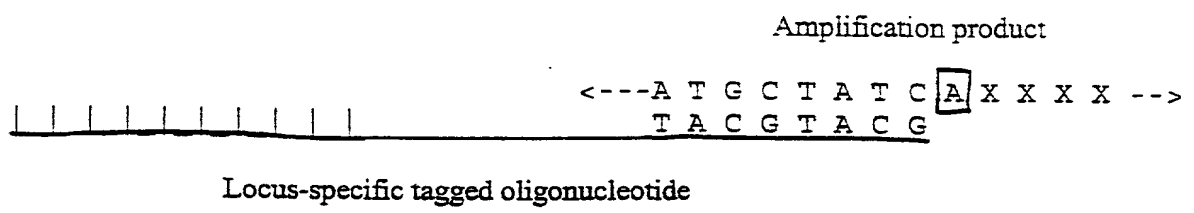


Fig. 3

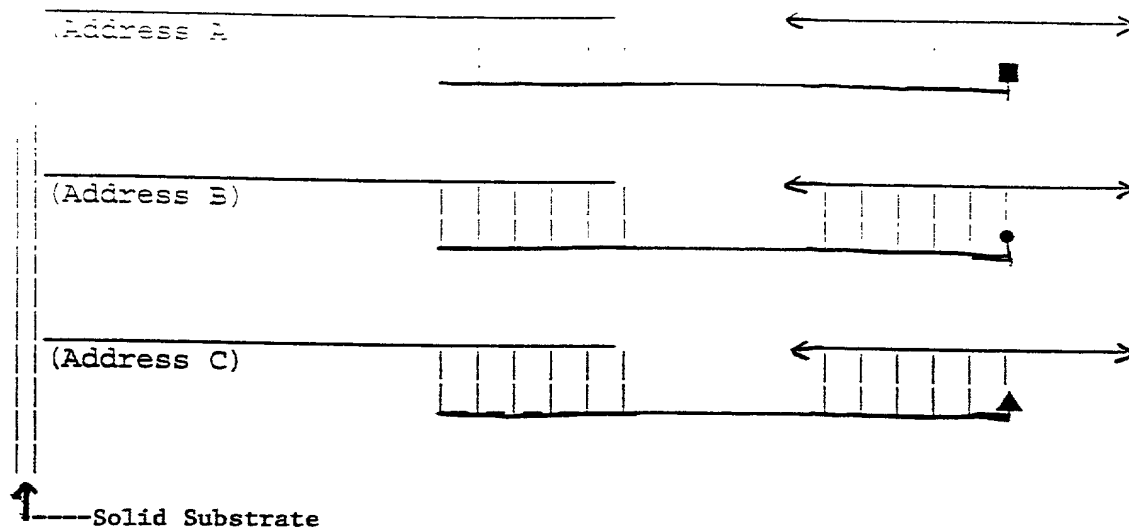


Fig. 5

Address A

(Address A)

(Address B)

(Address B)

(Address C)

(Address C)

Fig. 6

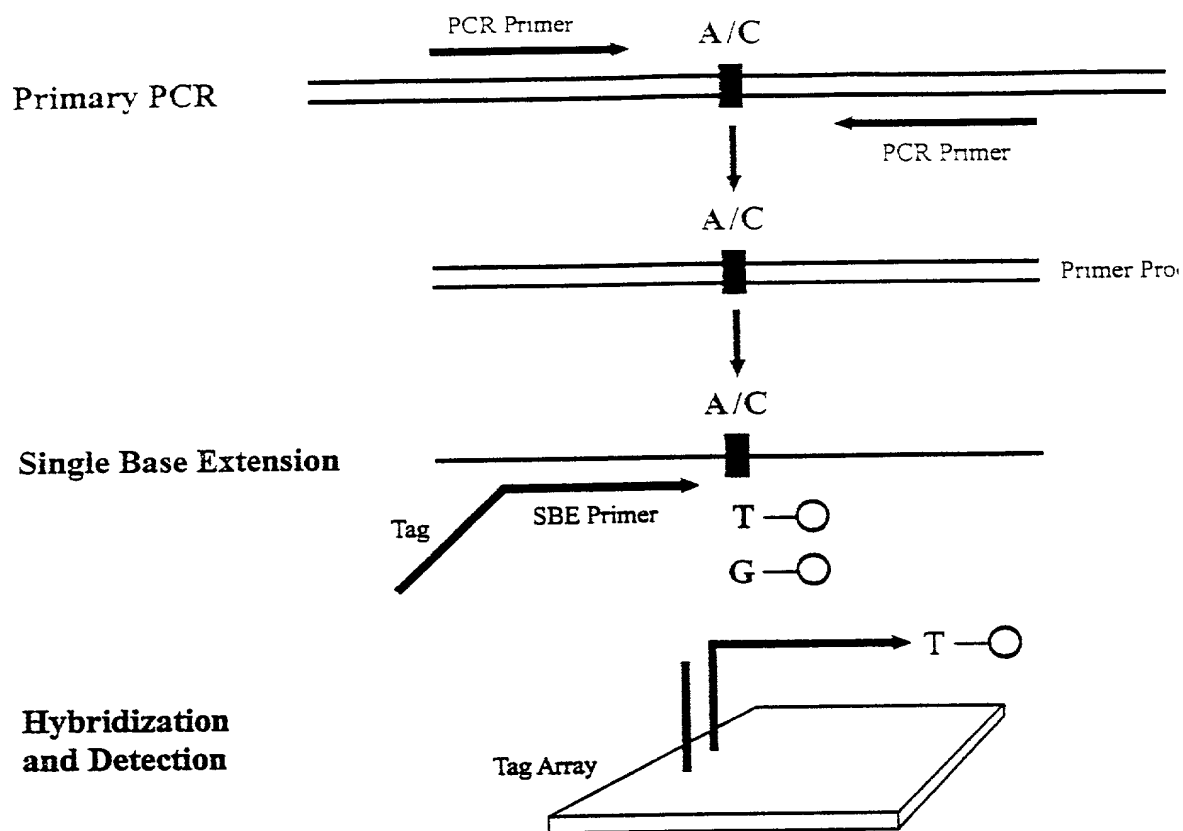


Fig. 7

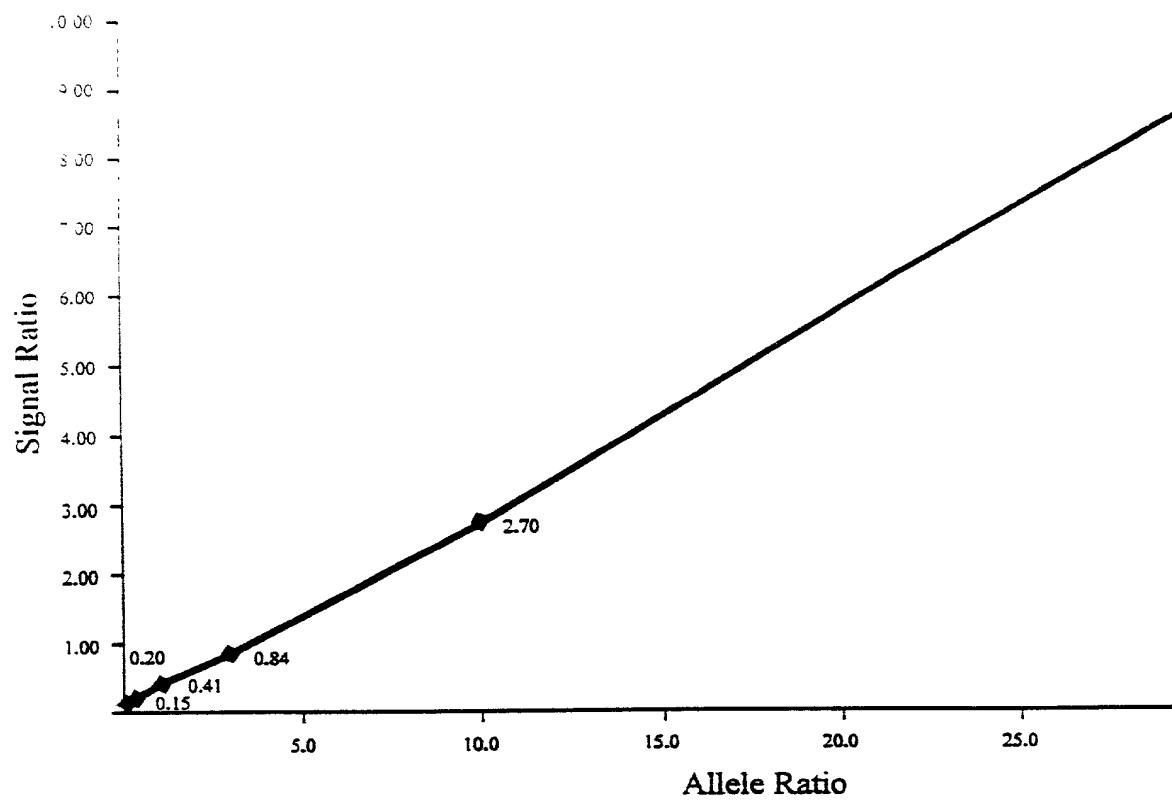


Fig. 8

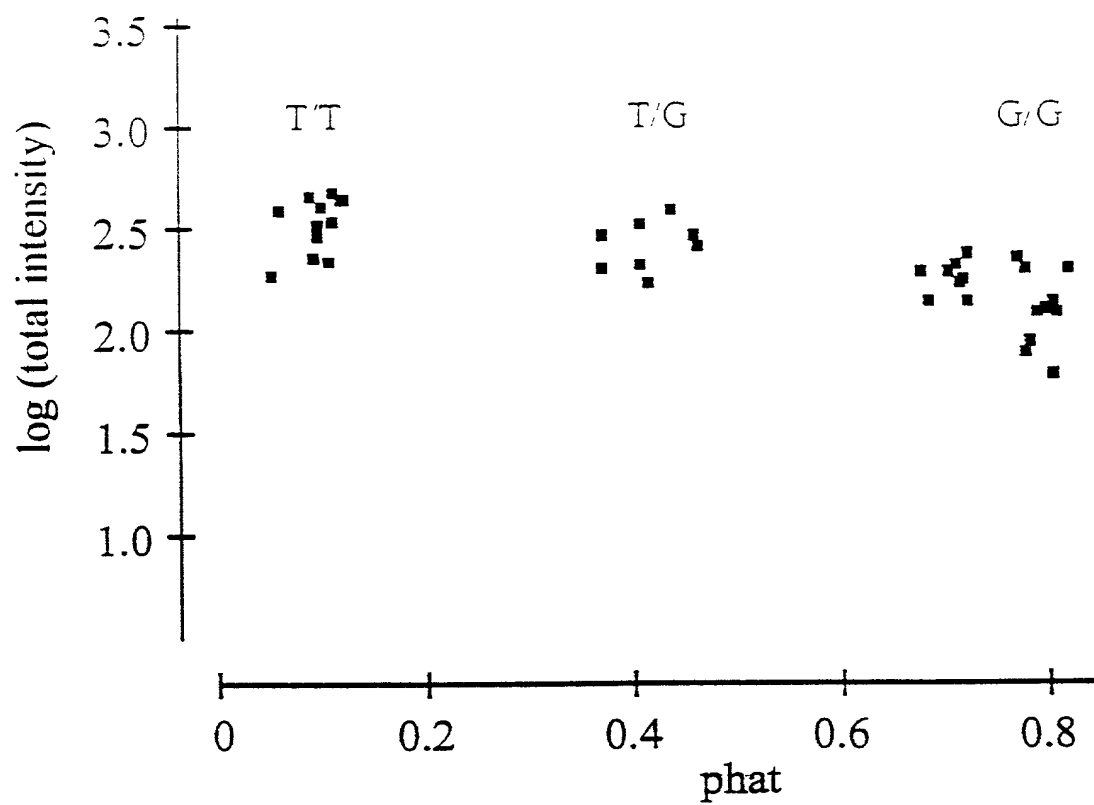


Fig. 9